

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 2, 2005, 18:55:16 ; Search time 803 Seconds
(without alignments)

6481.969 Million cell updates/sec

Title: US-10-649-433C-1

Perfect score: 2180

Sequence: 1 gtttagacgttatcatcg.....agtttagcggaattgatctg 6729

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2105692 seqs, 386760381 residues

Word size: 1

Total number of hits satisfying chosen parameters: 3967864

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-Q=/cn2_1/USFTO_spool_p/US10649433/runat_02082005_165025_1814/app_query.fasta_1.6919
-MODE=frame+ n2p.model -DEV=xlp
-DB=A Geneseq -QFWT=fastan -SUFFIX=oligo.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFWT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10649433 @CN 1 1 977 @runat 02082005_165025_1814 -NCPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THRSADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : A Geneseq_16Dec04:.*
1: geneseqp1980s:.*
2: geneseqp1990s:.*
3: geneseqp2000s:.*
4: geneseqp2001s:.*
5: geneseqp2002s:.*
6: geneseqp2003as:.*
7: geneseqp2003bs:.*
8: geneseqp2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	363	16.7	363	ABP51639	Abp51639 Escherich
2	360	16.5	360	AAR40926	Aar40926 Wild-type
3	360	16.5	360	AAW58898	AAW58898 Plasmid p
4	359	16.5	360	AAW80734	AAW80734 Amino aci
5	359	16.5	360	AAU79467	AAU79467 pNO23097
6	359	16.5	360	AAE15450	AAE15450 Lac repre
7	359	16.5	370	AAR40927	Aar40927 Cpg deple
8	359	16.5	372	AAR40927	Aar40927 Cpg deple
9	310	14.2	799	ABR43385	ABR43385 Nucleotid
10	309	14.2	309	ABR43374	ABR43374 E. coli g

11	309	14.2	332	4	AAU29371	Aau29371 Novel mar
12	309	14.2	332	6	ADA62006	Ada62006 E. coli g
13	306	14.0	329	2	AAR21083	Aar21083 GBP-delta
14	306	14.0	366	2	AAR21083	Aar21083 GBP-delta
15	296	13.6	1277	7	AAR52701	Aar52701 Humanid p
16	290	13.3	1293	7	ADC22811	Adc22811 Human G p
17	290	13.3	1293	7	ADH14284	Adh14284 Vector pc
18	290	13.3	1967	6	ABB82793	Abb82793 Amino aci
19	286	13.1	286	2	AAV06551	Aay06551 Escherich
20	286	13.1	872	7	ADK5686	Adk5686 E. coli cl
21	285	13.1	360	8	ADR16185	Adr16185 pET24b pl
22	284	13.0	360	2	AAV18028	Aay18028 E. coli l
23	284	13.0	360	6	ADA62016	Ada62016 E. coli m
24	284	13.0	360	6	ADA62017	Ada62017 E. coli m
25	284	13.0	360	6	ADA62011	Ada62011 E. coli m
26	284	13.0	360	6	ADA62013	Ada62013 E. coli m
27	284	13.0	360	6	ADA62001	Ada62001 E. coli w
28	284	13.0	360	6	ADA62012	Ada62012 E. coli m
29	284	13.0	360	6	ADA62014	Ada62014 E. coli m
30	284	13.0	360	6	ADA62015	Ada62015 E. coli m
31	284	13.0	1049	5	AAU75875	Aau75875 SREBP2-La
32	283	13.0	441	5	AAU75874	Aau75874 Notchi-La
33	283	13.0	1024	5	AAU75873	Aau75873 APP-Laci
34	272	12.5	360	6	ADA62018	Ada62018 E. coli m
35	244	11.2	319	8	ADI28562	Adi28562 LacIq rep
36	240	11.0	325	8	ADO76855	Adg76855 Parachyrc
37	240	11.0	1651	6	ABP70771	Abp70771 Insulin r
38	239	11.0	239	3	AAB22881	Aab22881 Enhanced
39	239	11.0	239	3	AAV79583	Aay79583 EYFP sign
40	239	11.0	239	5	ABG94443	Abg94443 Protease
41	239	11.0	239	6	AAG79830	Aag79830 Yellow fl
42	239	11.0	239	6	ADA38080	Ada38080 Aequorea
43	239	11.0	239	7	ADC18356	Adc18356 EYFP (not
44	239	11.0	239	7	ADF70401	Adf70401 Aequorea
45	239	11.0	239	7	ADG32368	Adg32368 Jellyfish

ALIGNMENTS

RESULT 1

ABP51639
ID ABP51639 standard; protein; 363 AA.

XX AC ABP51639;

XX DT 11-OCT-2002 (first entry)

XX DE Escherichia coli lacI repressor protein SEQ ID NO:29.

XX KW Corynebacterium glutamicum; transcriptional regulation; gene expression;
amino acid biosynthesis.

XX OS Escherichia coli.

XX PN WO200240679-A2.

XX PD 23-MAY-2002.

XX PF 15-NOV-2001; 2001WO-US043096.

XX PR 15-NOV-2000; 2000US-0248219P.

XX PA (RAYA/) RAYAPATI P J.

XX PA (CRAF/) CRAFTON C M.

XX PI Rayapati PJ, Crafton CM;

XX DR WPI; 2002-575217/61.

XX DR N-PSDB; ABQ73199.

XX PT Novel polynucleotides from Corynebacterium glutamicum useful for inducing
and regulating expression of genes, including those that are involved in
amino acid biosynthesis, in bacterial cells.

XX Disclosure; Page 107-109; 112pp; English.

XX The present invention describes Corynebacterium glutamicum

CC transcriptional regulatory region polynucleotide sequences (I). ABQ73173

CC to ABQ73194 represent the C. glutamicum transcriptional regulatory

CC regions pta, aceA, aceB, adh, aldB, poxB, ldh, amyB, malZ, bglX, gam,

CC glgX, hisD, pyrR, purD, hrcA, htpX, dnaK, ctc, grpE, clpB, and narX,

CC respectively. (I) can be used for producing a vector. (I) can also be

CC used for producing a transformed Corynebacterium sp. host cell, which is

CC useful for producing a biosynthetic product. (I) is useful for producing

CC a Corynebacterium sp. host cell, which is useful for producing an amino

CC acid. (I) is useful for regulating and enhancing the production of a

CC variety of products in host cells, including amino acids such as lysine,

CC purine nucleotides such as inosinic acid, and heterologous polypeptides.

CC The present sequence represents an E. coli lacI repressor protein from

CC the present invention

XX

SQ Sequence 363 AA;

Alignment Scores:

Pred. No.: 1.55e-303 Length: 363

Score: 363.00 Matches: 363

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 16.65% Indels: 0

DB: 5 Gaps: 0

US-10-649-433C-1 (1-6729) x ABP51639 (1-363)

QY 5599 GTGGTGAATGTGAACACAGTAACTTATACAGATGTCGACAGTAGTGCCTCTTAT 5658

DB 1 ValValAsnValLysProValThrLeuIysArgValAlaGluTyrAlaGlyValSerIyr 20

QY 5659 CAGACCGTTTCCCGCGTGGTGAACACAGCCAGCCACGTTTCTGCGAAACCGGGAAAAA 5718

DB 21 GlnThrValSerArgValValAsnGlnAlaSerHisValSerAlaLysThrArgGluLys 40

QY 5719 GTGGAGCGCGATGGCGAGCTGAATTACATTCCTCCACCGCGTGGCAACACTGGCG 5778

DB 41 ValGluAlaAlaMetAlaGluLeuAsnTyrIleProAsnArgValAlaGlnGlnLeuAla 60

QY 5779 GGCACACAGTCTGCTGATGGCGTGCACCTCCAGTCTGGCCCTGCACGCGCGTCG 5838

DB 61 GlyLysGlnSerLeuLeuIleGlyValAlaThrSerSerLeuAlaLeuHisAlaProSer 80

QY 5839 CAAATTGTCGGCGGATTAATCTCCGCGCGATCAACTGGGTGCGAGCGTGGTGTGTCG 5898

DB 81 GlnIleValAlaAlaIleLysSerArgAlaAspGlnLeuGlyAlaSerValValSer 100

QY 5899 ATGTAGAACGAGCGGTGCGAGCCCTGTAAGCGCGGTGCACAACTCTTCGCGCAA 5958

DB 101 MetValGluArgSerGlyValGluAlaCysLysAlaAlaValHisAsnLeuLeuAlaGln 120

QY 5959 CGCGTCAGTGGCGTCAATTAACATATCCGTGGATGACACAGGATGCCATTTGCTGGAA 6018

DB 121 ArgValSerGlyLeuIleAsnTyrProLeuAspAspGlnAspAlaIleAlaValGlu 140

QY 6019 GCTGCTGCACATAATGTTCCGCGGTTATTTCTTGATGTCTCTGACACACACCCATCAAC 6078

DB 141 AlaAlaCysThrAsnValProAlaLeuPheLeuAspValSerAspGlnThrProIleAsn 160

QY 6079 AGTATTATTTCTCCATGAGACCGGTACGCGACTGGCGGTGGACATCTGTCGATTG 6138

DB 161 SerIleIlePheSerHisGluAspGlyThrArgLeuGlyValGluHisLeuValAlaLeu 180

QY 6139 GGTACACAGCAATCGCGTGTAGCGGCCCATTAAGTTCCTCGCGCGCTCTGGGT 6198

DB 181 GlyHisGlnGlnIleAlaLeuLeuAlaGlyProLeuSerSerValSerAlaArgLeuArg 200

QY 6199 CTGGTGGTGGCATAAATATCTCACTCGCAATCAAAATTCAGCCGATAGCGGAACGGAA 6258

DB 201 LeuAlaGlyTyrHisLysTyrLeuThrArgAsnGlnIleGlnProIleAlaGluArgGlu 220

QY 6259 GCGACTGGAGTCCATGCTCCGGTTTCAACAAACCATGCAATGCTGAATGAGGCGATC 6318

DB 221 GlyAspTrpSerAlaMetSerGlyPheGlnGlnThrMetGlnMetLeuAsnGluGlyIle 240

QY 6319 GTTCCCACTGGATGCTGGTTGGCCAAACGATCAGATGGCGCTGGCGCAATGCGCCCAT 6378

DB 241 ValProThrAlaMetLeuValAlaAsnAspGlnMetAlaLeuGlyAlaMetArgAlaIle 260

QY 6379 ACCGAGTCCGGCTGCGGCTGGTGGCGGATATCTCGTAGTGGGATACGATACCGAA 6438

DB 261 ThrGluSerGlyLeuArgValGlyAlaAspIleSerValValGlyTyrAspAspThrGlu 280

QY 6439 GACAGCTCAGTATATATCCCGCTCAACACCATCAACAGGATTTTCGCTGCTGGGG 6498

DB 281 AspSerSerCysTyrIleProProSerThrThrIleLysGlnAspPheArgLeuLeuGly 300

QY 6499 CAAACACGCTGACCGCTTCTCTCAGGGCCAGCGGTGAAGGGCAATCAG 6558

DB 301 GlnThrSerValAspArgLeuLeuGlnSerGlnGlyGlnAlaValLysGlyAsnGln 320

QY 6559 CTGTTGCCGCTCTCACTGGTGAAGAAACCAACCCCTGGCGGCCCAATACGCAACCGCC 6618

DB 321 LeuLeuProValSerLeuValLysArgLysThrThrLeuAlaProAsnThrGlnThrAla 340

QY 6619 TCTCCCGCGCGTGGCGGATTCATTAATGACGCTGGCAGCAGAGTTTCCGCTGCTGAA 6678

DB 341 SerProArgAlaLeuAlaAspSerLeuMetGlnLeuAlaArgGlnValSerArgLeuGlu 360

QY 6679 AGCGGCGAG 6687

DB 361 SerGlyGln 363

RESULT 2

AAR40926

ID AAR40926 standard; protein; 360 AA.

XX

AC AAR40926;

XX 25-MAR-2003 (revised)

DT 09-JAN-2003 (revised)

DT 22-FEB-1994 (first entry)

XX

DE Wild-type lacI.

XX Bacteria; CpG doublet; vertebrate; mammalian; methylation; marker;

KW reporter; transgene; mutagen; carcinogen; mutation; assay;

XX chemical agent.

OS Bacteriophage lambda.

XX WO9317123-A1.

XX

PD 02-SEP-1993.

XX

XX 26-FEB-1993; 93WO-US001676.

XX

PR 27-FEB-1992; 92US-00842664.

XX

PA (OHIS) UNIV OHIO STATE.

XX

PI Stambrook PJ;

XX

DR WPI; 1993-288426/36.

DR N-PSDB; AAQ48615.

XX

PT Modified mutagenicity test - comprises assaying transgenic animals for

PT genetic mutations, utilises native amount of CpG nucleotide(s).

XX

PS Disclosure; Page 52-53; 79pp; English.

XX

CC Bacterial genes exhibit a much higher frequency of occurrence of the CpG

CC doublet than do vertebrate genes. As a result, a bacterial gene

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 2, 2005, 19:39:42 ; Search time 159.5 Seconds
(without alignments)

6298.603 Million cell updates/sec

Title: US-10-649-433C-1

Perfect score: 2180

Sequence: 1 gtttagacgttatcatcg.....agttagcgcgaattgatctg 6729

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 513545 seqs, 74649064 residues

Word size: 1

Total number of hits satisfying chosen parameters: 903960

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-Q/cgn2_1/USPTO_spool_p/US10649433/runat_02082005_165026_1844/app_query.fasta_1.6919
-DB=issued_Patents_AA -Qfmt=fastan -SUFFIX=Oligo.ra1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10649433 @CNC 1.1 178 @runat_02082005_165026_1844 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/prodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/prodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/prodata/1/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/prodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	360	16.5	360	5	PCT-US93-01676A-2
2	359	16.5	370	5	PCT-US93-01676A-4
3	290	13.3	1293	4	US-09-170-496D-292
4	290	13.3	1293	4	US-09-364-425B-57
5	286	13.1	286	4	US-09-555-510B-9
6	286	13.1	286	4	US-10-231-013-9
7	285	13.1	360	3	US-09-025-769B-293
8	285	13.1	360	4	US-09-490-070A-293
9	285	13.1	360	4	US-09-490-153-293
10	285	13.1	360	4	US-09-490-324-293
11	284	13.0	1049	3	US-09-522-666-6
12	283	13.0	441	3	US-09-522-666-4

13	283	13.0	1024	3	US-09-522-666-2	Sequence 2, Appli
14	244	11.2	319	3	US-09-044-796A-13	Sequence 13, Appl
15	244	11.2	319	4	US-09-725-460A-13	Sequence 13, Appl
16	239	11.0	239	4	US-09-513-783A-44	Sequence 44, Appl
17	239	11.0	239	4	US-09-430-656-44	Sequence 44, Appl
18	239	11.0	541	4	US-09-513-783A-34	Sequence 34, Appl
19	239	11.0	541	4	US-09-430-656-34	Sequence 34, Appl
20	239	11.0	812	4	US-09-513-783A-4	Sequence 4, Appli
21	239	11.0	812	4	US-09-513-783A-6	Sequence 6, Appli
22	239	11.0	812	4	US-09-430-656-4	Sequence 4, Appli
23	239	11.0	812	4	US-09-430-656-6	Sequence 6, Appli
24	239	11.0	1610	4	US-09-513-783A-22	Sequence 22, Appl
25	239	11.0	1610	4	US-09-430-656-22	Sequence 22, Appl
26	238	10.9	1056	4	US-09-513-783A-32	Sequence 32, Appl
27	238	10.9	1056	4	US-09-430-656-32	Sequence 32, Appl
28	204	9.4	286	1	US-07-721-775A-2	Sequence 2, Appli
29	204	9.4	286	1	US-08-339-658-2	Sequence 2, Appli
30	204	9.4	286	3	US-09-263-933-7	Sequence 7, Appli
31	204	9.4	286	3	US-09-263-933-14	Sequence 14, Appl
32	204	9.4	286	3	US-09-263-933-21	Sequence 21, Appl
33	204	9.4	286	3	US-09-025-769B-265	Sequence 265, App
34	204	9.4	286	3	US-09-025-769B-362	Sequence 362, App
35	204	9.4	286	4	US-09-919-901-7	Sequence 7, Appli
36	204	9.4	286	4	US-09-919-901-14	Sequence 14, Appl
37	204	9.4	286	4	US-09-919-901-21	Sequence 21, Appl
38	204	9.4	286	4	US-09-490-070A-265	Sequence 265, App
39	204	9.4	286	4	US-09-490-070A-362	Sequence 362, App
40	204	9.4	286	4	US-09-490-153-265	Sequence 265, App
41	204	9.4	286	4	US-09-490-153-362	Sequence 362, App
42	204	9.4	286	4	US-10-191-966-7	Sequence 7, Appli
43	204	9.4	286	4	US-10-191-966-14	Sequence 14, Appl
44	204	9.4	286	4	US-10-191-966-21	Sequence 21, Appl
45	204	9.4	286	4	US-09-490-324-265	Sequence 265, App

ALIGNMENTS

RESULT 1
PCT-US93-01676A-2
; Sequence 2, Application PC/TUS9301676A
; GENERAL INFORMATION:
; APPLICANT: Scambrook, Peter J.
; TITLE OF INVENTION: MUTAGENICITY TESTING USING REPORTER
; TITLE OF INVENTION: GENES WITH MODIFIED METHYLATION FREQUENCIES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/01676A
; FILING DATE: 19930226
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 07/842,644
; FILING DATE: 02-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Cooper, Iver P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: STAMBOOK 1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 360 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-01676A-2

Alignment Scores:

Pred. No.: 4,566-298 Length: 360
Score: 360.00 Matches: 360
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 16.51% Indels: 0
DB: 5 Gaps: 0

US-10-649-433C-1 (1-6729) x PCT-US93-01676A-2 (1-360)

QY 5608 GTGAACACCAAGTATTATAGATCTCGCAGATATGCGGTGTCTCTTATCAGACCGTT 5667
DB 1 VallysProValThrLeuTyrAspValAlaGluTyrAlaGlyValSerTyrGlnThrVal 20
QY 5668 TCCCGGTGGTAACACGCGCAGCCACGCTTCTGCGAAACGCGGAAAAAGTGAACGG 5727
DB 21 SerArgValValAsnGlnAlaSerHisValSerAlaLysThrArgGluLysValGluAla 40
QY 5728 CGCATGGCGGAGCTGAATTACATTCCTCCCAACCGCTGGCACAACACTGCGCGGCAACAG 5787
DB 41 AlaMetAlaGluLeuAsnTyrIleProAsnArgValAlaGlnGlnLeuAlaGlyLysGln 60
QY 5788 TCGTGTGATGGCGTTGCCACCTCCAGTCTGGCCCTGCACGCGCGCTCGCAAAATGTC 5847
DB 61 SerLeuLeuIleGlyValAlaThrSerSerLeuAlaLeuHisAlaProSerGlnIleVal 80
QY 5848 GCGCGATTAATCTCGCGCCGATCAACTGGGTGCCAGCGTGGTGCATGATGTAGAA 5907
DB 81 AlaAlaIleLysSerArgAlaAspGlnLeuGlyAlaSerValValValSerMetValGlu 100
QY 5908 CGAGCGCGCTCGAGCTGTAAAGCGCGGTGCACAACTCTCTCGCGCAACGGTCACT 5967
DB 101 ArgSerGlyValGluAlaCysLysAlaAlaValHisAsnLeuLeuAlaGlnArgValSer 120
QY 5968 GGGCTGATCAATTAATCCGCTGATGACAGATGCAATTCCTGTGGAAGTCCCTGCG 6027
DB 121 GlyLeuIleIleAsnTyrProLeuAspAspGlnAspAlaIleAlaValGluAlaCys 140
QY 6028 ACTAATGTCGCGCTTATTTCTGATGCTCTGACCAAGACCACTCAACAGATTAAT 6087
DB 141 ThrAsnValProAlaLeuPheLeuAspValSerAspGlnThrProIleAsnSerIle 160
QY 6088 TTCTCCCATGAAGCGTACCGACTGGCGTGGAGCATCTGTCGCATTCGGTCCACAG 6147
DB 161 PheSerHisGluAspGlyThrArgLeuGlyValGluHisLeuValAlaLeuGlyHisGln 180
QY 6148 CAATTCGCGCTGTAGCGGCGCCATTAAAGTTCTGTCGCGCGCTCTGCTCTGCTGCG 6207
DB 181 GlnIleAlaLeuLeuAlaGlyProLeuSerSerValSerAlaArgLeuArgLeuAlaGly 200
QY 6208 TGGCAATAATATCTCACTCGCAATCAAAATTCAGCCGATAGCGGAACGGGAAGCGACTGG 6267
DB 201 TrpHisLysTyrLeuThrArgAsnGlnIleGlnProIleAlaGluArgGluGlyAspTrp 220
QY 6268 AGTGCATGTCGGGTTTTCACAAACCATGCAATGCTGAATGAGGGGCATCGTTCCTCACT 6327
DB 221 SerAlaMetSerGlyPheGlnThrMetGlnMetLeuAsnGluGlyIleValProThr 240
QY 6328 CGCATGCTGTTGCCAACGATCAGATGGCGTGGCGCAATCGCGCCATACCGAGTCC 6387
DB 241 AlaMetLeuValAlaAsnAspGlnMetAlaLeuGlyAlaMetArgAlaIleThrGluSer 260
QY 6388 GGGCTGCGGTTGGTCCGATATCTCGGTAGTGGGATACGAGATACCGAAGACAGCTCA 6447
DB 261 GlyLeuArgValGlyAlaAspIleSerValValGlyTyrAspAspThrGluAspSerSer 280

QY 6448 TGTATATCCCGCGCTCAACACCATCAACAGAGATTTTCCTGCTGGGCAACACGAGC 6507
DB 281 CystyrIleProSerThrThrIleLysGlnAspPheArgLeuLeuGlyGlnThrSer 300
QY 6508 GTGACCGCTTGTCTGCAACTCTCTCAGGCGCAGCGGTGAAGGGCAATCAGCTGTGCC 6567
DB 301 ValAspArgLeuLeuGlnLeuSerGlnGlyGlnAlaValLysGlyAsnGlnLeuLeuPro 320
QY 6568 GTCTACCTGGTGAAGAAAAACACCTCTGCGGCGCCCAATAGCAAAACCGCTCTCCGCGC 6627
DB 321 ValSerLeuValLysArgLysThrThrLeuAlaProAsnThrGlnThrAlaSerProArg 340
QY 6628 GCGTTGGCCGATTCATTAATCAGCTGCGCAGCAGAGTTTCCCGACTCGAAAGCGGCGAG 6687
DB 341 AlaLeuAlaAspSerLeuMetGlnLeuAlaArgGlnValSerArgLeuGluSerGlyGln 360

RESULT 2
PCT-US93-01676A-4
; Sequence 4, Application PC/TUS9301676A
; GENERAL INFORMATION:
; APPLICANT: Stambrook, Peter J.
; TITLE OF INVENTION: MUTAGENICITY TESTING USING REPORTER
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/01676A
; FILING DATE: 19930226
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/842,644
; FILING DATE: 02-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Cooper, Iver P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: STAMBROOK 1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 370 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-01676A-4

Alignment Scores:
Pred. No.: 3,23e-297 Length: 370
Score: 359.00 Matches: 359
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 16.47% Indels: 0
DB: 5 Gaps: 0

US-10-649-433C-1 (1-6729) x PCT-US93-01676A-4 (1-370)

QY 5611 AAACAGTAAGTTATACGATGTCGAGATATGCGGTGTCTCTTATCAGACCGTTTCC 5670
DB 2 LysProValThrLeuTyrAspValAlaGluTyrAlaGlyValSerTyrGlnThrValSer 21

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: August 2, 2005, 20:07:07 ; Search time 801 Seconds
(without alignments)
6545.998 Million cell updates/sec

Title: US-10-649-433C-1
Perfect score: 2180
Sequence: 1 gtttgacagcttatcatcg.....agttagcggaattgatctg 6729

Scoring table:
Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1745140 seqs, 389608008 residues

Word size: 1

Total number of hits satisfying chosen parameters: 3345576

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-Q=/cgn2_1/USPTO_spool_p/US10649433/runat_02082005_165027_1863/app_query.fasta_1.6919
-DB=Published Applications AA -QWITS=fastan -SUFFIX=Oligo.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=quality -THR MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USRP=US10649433 @CNG 1.1 964 @runat_02082005_165027_1863
-NCPUL=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published Applications AA:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pcp.*
2: /cgn2_6/ptodata/1/pubpaa/US06_PCT_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pcp.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pcp.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pcp.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pcp.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pcp.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pcp.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pcp.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pcp.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pcp.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pcp.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pcp.*
18: /cgn2_6/ptodata/1/pubpaa/US10F_NEW_PUB.pcp.*
19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pcp.*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pcp.*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pcp.*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	363	16.7	363	10	US-09-987-763-29 Sequence 29, Appl
2	359	16.5	372	16	US-10-469-881-3 Sequence 3, Appl
3	309	14.2	309	16	US-10-840-796-1 Sequence 1, Appl
4	309	14.2	332	13	US-10-197-053-6 Sequence 6, Appl
5	309	14.2	332	17	US-10-893-671-82 Sequence 82, Appl
6	290	13.3	1293	14	US-10-251-385-292 Sequence 292, Appl
7	290	13.3	1293	18	US-10-668-035-57 Sequence 57, Appl
8	290	13.3	1967	16	US-10-477-044-1 Sequence 1, Appl
9	286	13.1	286	14	US-10-231-013-9 Sequence 9, Appl
10	284	13.0	286	16	US-10-842-534-9 Sequence 9, Appl
11	284	13.0	360	13	US-10-197-053-1 Sequence 1, Appl
12	244	11.2	319	9	US-09-725-460A-13 Sequence 13, Appl
13	244	11.2	319	20	US-11-004-853-15 Sequence 15, Appl
14	239	11.0	239	14	US-10-162-593-3 Sequence 3, Appl
15	239	11.0	239	14	US-10-100-957A-44 Sequence 44, Appl
16	239	11.0	239	15	US-10-370-570-5 Sequence 5, Appl
17	239	11.0	239	15	US-10-442-148A-5 Sequence 5, Appl
18	239	11.0	239	16	US-10-724-178-4 Sequence 4, Appl
19	239	11.0	239	17	US-10-763-976A-42 Sequence 42, Appl
20	239	11.0	239	17	US-10-763-976A-55 Sequence 55, Appl
21	239	11.0	239	17	US-10-505-486-24 Sequence 24, Appl
22	239	11.0	253	15	US-10-442-148A-6 Sequence 6, Appl
23	239	11.0	541	14	US-10-100-957A-34 Sequence 34, Appl
24	239	11.0	643	16	US-10-838-770-4 Sequence 4, Appl
25	239	11.0	812	14	US-10-100-957A-4 Sequence 4, Appl
26	239	11.0	812	14	US-10-100-957A-6 Sequence 6, Appl
27	239	11.0	1610	14	US-10-100-957A-22 Sequence 22, Appl
28	238	10.9	1056	14	US-10-100-957A-32 Sequence 32, Appl
29	237	10.9	238	17	US-10-505-486-5 Sequence 5, Appl
30	237	10.9	387	11	US-09-927-876-87 Sequence 87, Appl
31	237	10.9	387	15	US-10-360-149-87 Sequence 87, Appl
32	237	10.9	387	15	US-10-457-047-87 Sequence 87, Appl
33	237	10.9	387	17	US-10-811-081-87 Sequence 87, Appl
34	237	10.9	387	17	US-10-913-944-87 Sequence 87, Appl
35	237	10.9	545	14	US-10-214-932-52 Sequence 52, Appl
36	237	10.9	548	14	US-10-214-932-76 Sequence 76, Appl
37	237	10.9	1192	14	US-10-214-932-54 Sequence 54, Appl
38	225	10.3	360	13	US-10-197-053-2 Sequence 2, Appl
39	204	9.4	263	15	US-10-668-778-2 Sequence 2, Appl
40	204	9.4	264	17	US-10-877-952-18 Sequence 18, Appl
41	204	9.4	286	10	US-09-919-901-7 Sequence 7, Appl
42	204	9.4	286	10	US-09-919-901-14 Sequence 14, Appl
43	204	9.4	286	10	US-09-919-901-21 Sequence 21, Appl
44	204	9.4	286	11	US-09-837-306-354 Sequence 354, App
45	204	9.4	286	14	US-10-191-966-7 Sequence 7, Appl

ALIGNMENTS

RESULT 1
US-09-987-763-29
; Sequence 29, Application US/09987763
; Publication No. US20030017553A1
; GENERAL INFORMATION:
; APPLICANT: Rayapati, P. John
; APPLICANT: Crafton, Corey M.
; TITLE OF INVENTION: Nucleotide Sequences for Transcriptional Regulation in
; TITLE OF INVENTION: Corynebacterium glutamicum
; FILE REFERENCE: 1533-194002
; CURRENT APPLICATION NUMBER: US/09/987,763
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: 60/248,219
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 29
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Escherichia coli

US-09-987-763-29

Alignment Scores:

Pred. No.: 0 Length: 363
Score: 363.00 Matches: 363
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 16.65% Indels: 0
DB: 10 Gaps: 0

US-10-649-433C-1 (1-6729) x US-09-987-763-29 (1-363)

QY 5599 GTGGTGAATGTCARACAGTACGTTATACGATGTCGACAGTATGCCGGTCTCTCTTAT 5658
Db 1 ValValAenValLysProValThrLeuTyrAspValAlaGluTyrAlaGlyValSerTyr 20
QY 5659 CAGACCGTTTCCCGCGTGGTGAACACGACGCCACCGTTTCTGCGAAAAACGGGAAAAA 5718
Db 21 GlnThrValSerArgValValAenGlnAlaSerHisValSerAlaLysThrArgGluLys 40
QY 5719 GTGGAAGCGCGATGGCGAGCTGAATTACATCCAAACCGCGTGGCAACAACATCGCG 5778
Db 41 ValGluAlaAlaMetAlaGluLeuAenTyrIleProAsnArgValAlaGlnGlnLeuAla 60
QY 5779 GGCRAACAGTCGTTGCTGATTGGCGTGCACCTCCAGTCTGGCCCTGCACGCGCGTCG 5838
Db 61 GlyLysGlnSerLeuLeuIleGlyValAlaThrSerSerLeuAlaLeuHisAlaProSer 80
QY 5839 CAAATTGTCCGCGGATTAAATCTCCGCGCATCAACTGGGTGGTGGTGGTGGTGGTGG 5898
Db 81 GlnIleValAlaAlaIleLysSerArgAlaAspGlnLeuGlyAlaSerValValSer 100
QY 5899 ATGGTAGAACGAGCGCGTGAAGCGGTGTAAAGCGCGGTGCACAACTCTCTCCGCAA 5958
Db 101 MetValGluArgSerGlyValGluAlaCysLysAlaAlaValHisAsnLeuLeuAlaGln 120
QY 5959 CGCGTCAGTGGCTCATTAACATATCGCTGGATGACCGAGTGCCTGCTGTGGAA 6018
Db 121 ArgValSerGlyLeuIleLeuAsnTyrProLeuAspGlnAspAlaIleAlaValGlu 140
QY 6019 GCTGCTGCATTAATGTTCCGCGTATTCTTGATGTCTCTGACGACACCCATCAAC 6078
Db 141 AlaAlaCysThrAsnValProAlaLeuPheLeuAspValSerAspGlnThrProIleAsn 160
QY 6079 AGTATTATTCTCCATGAACACGCTAGCGACCTGGCGGTGGAGCATCTGCTCCATTG 6138
Db 161 SerIleIlePheSerHisGluAspGlyThrArgLeuGlyValGluHisLeuValAlaLeu 180
QY 6139 GGTACACGACAAATCGCGTGTAGCGGCCCATTAAGTTCTGCTCGCGGTCTGCGT 6198
Db 181 GlyHisGlnGlnIleAlaLeuLeuAlaGlyProLeuSerSerValSerAlaArgLeuArg 200
QY 6199 CTGGCTGGCTGGCATAAATATCTCACTCGCAATCAAATTCAGCCGATAGCGGAACGGAA 6258
Db 201 LeuAlaGlyTyrHisLysTyrLeuThrArgAsnGlnIleGlnProIleAlaGluArgGlu 220
QY 6259 GCGAGTCGAGTGCATGCTCCGGTTTCAAACCAATGCAAAATCTGAATGAGGCGATC 6318
Db 221 GlyAspTyrSerAlaMetSerGlyPheGlnGlnThrMetGlnMetLeuAsnGluGlyIle 240
QY 6319 GTTCCACTCGCATGCTGGTTCGCAACATCATAGTGGCGCTGGGCGCAATGCGCCCAAT 6378
Db 241 ValProThrAlaMetLeuValAlaAsnAspGlnMetAlaLeuGlyAlaMetArgAlaIle 260
QY 6379 ACCGAGTCGGGCTCGCGTGGTGGGATATCTCGGTAGTGGGATACGACGATACCGAA 6438
Db 261 ThrGluSerGlyLeuArgValGlyAlaAspIleSerValValGlyTyrAspAspThrGlu 280
QY 6439 GACAGTCATGTTATATCCGCGGTCAACCAACCATCAACAGATTTCGCTGCTGGGG 6498
Db 281 AspSerSerCysTyrIleProProSerThrThrIleLysGlnAspPheArgLeuLeuGly 300
QY 6499 CAAACACGCGTGGACCGCTTGTCGAATCTCTCAGGGCCAGCGCGGTGAAGGGCAATCAG 6558

Db 301 GlnThrSerValAspArgLeuLeuGlnLeuSerGlnGlyGlnAlaValLysGlyAsnGln 320
QY 6559 CTGTTGCCCGTCTCACTGGTCAAAAAGAAAAACCACTGGCGCCCAATACGCAACCGCC 6618
Db 321 LeuLeuProValSerLeuValLysArgLysThrThrLeuAlaProAsnThrGlnThrAla 340
QY 6619 TCTCCCGCGCGTGGCGCGATTCAATATGAGCTGGGACGACAGGTTTCCGACTGGA 6678
Db 341 SerProArgAlaLeuAlaAspSerLeuMetGlnLeuAlaArgGlnValSerArgLeuGlu 360
QY 6679 AGCGGGCAG 6687
Db 361 SerGlyGln 363
RESULT 2
US-10-469-881-3
; Sequence 3, Application US/10469881
; Publication No. US20040171824A1
; GENERAL INFORMATION:
; APPLICANT: The University of Virginia Patent Foundation
; APPLICANT: Scrable, Heidi J
; APPLICANT: Cronin, Carolyn A
; TITLE OF INVENTION: A Lac Operator-Repressor System
; FILE REFERENCE: 00663-02
; CURRENT APPLICATION NUMBER: US/10/469,881
; CURRENT FILING DATE: 2003-09-04
; PRIOR APPLICATION NUMBER: 60/281,322
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/273,480
; PRIOR FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 3
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-469-881-3
Alignment Scores:
Pred. No.: 0 Length: 372
Score: 359.00 Matches: 359
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 16.47% Indels: 0
DB: 16 Gaps: 0
US-10-649-433C-1 (1-6729) x US-10-469-881-3 (1-372)
QY 5611 AAACCACTAAGTTATAGATGTCGACAGTATGCCGGTGTCTCTTATCAGACCGTTTCC 5670
Db 2 LysProValThrLeuTyrAspValAlaGluTyrAlaGlyValSerTyrGlnThrValSer 21
QY 5671 CGCGTGGTGAACACGACGCCACCGTTTCTGCAAAAACCGGAAAAAGTGAAGCGGG 5730
Db 22 ArgValValAenGlnAlaSerHisValSerAlaLysThrArgGluLysValGluAla 41
QY 5731 ATGCGGAGCTGAATTACATTCCTCCAAACCGCGTGGCAACAACCTGCGCGGCAACAGTCG 5790
Db 42 MetAlaGluLeuAsnTyrIleProAsnArgValAlaGlnGlnLeuAlaGlyLysGlnSer 61
QY 5791 TTGCTGATTGGCTTGGCCACCTCCAGTCTGCCCTGACGCGCGTCCCAATGTCGCG 5850
Db 62 LeuLeuIleGlyValAlaThrSerSerLeuAlaLeuHisAlaProSerGlnIleValAla 81
QY 5851 GCGATTAAATCTCGCGCGCATCAACTGGGTGCCAGCGTGGTGGTTCGATGCTAGAACGA 5910
Db 82 AlaIleLysSerArgAlaAspGlnLeuGlyAlaSerValValValSerMetValGluArg 101
QY 5911 AGCGCGTCGAAGCTGTAAAGCGCGGTGACCAATCTTCTCGCGCAACGCTCAGTGGG 5970
Db 102 SerGlyValGluAlaCysLysAlaAlaValHisAsnLeuLeuAlaGlnArgValSerGly 121

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 2005, 19:06:10 ; Search time 27848 Seconds
(without alignments)
11708.387 Million cell updates/sec

Title: US-10-649-433C-1

Perfect score: 6729

Sequence: 1 gtttgacagcttatcatcga.....agttagcggaattgatctg 6729

Scoring table: OLIGO NUC

Gapop_60.0 , Gapext 60.0

Searched: 4708233 seqs, 24227607955 residues

Word size : 0

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sta.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3855	57.3	4806	12 AF050464	AF050464 Expressio
2	3804	56.5	5013	6 AR264513	AR264513 Sequence
3	3753	55.8	4176	6 A13038	A13038 Artificial
4	3753	55.8	4176	12 SYNTRC99A	M22744 Cloning vec
5	3753	55.8	4239	6 A29289	A29289 pSEC-Bp1 co
6	3753	55.8	5271	12 CVU89936	U89936 Cloning vec
7	3752	55.8	4476	6 A102989	AR102989 Sequence
8	3752	55.8	4476	6 AR172783	AR172783 Sequence
9	3703	55.0	4214	6 BD273391	BD273391 P450/acet
10	3703	55.0	4214	6 AX027488	AX027488 Sequence
11	3702	55.0	4176	12 XXU13872	U13872 pirc99a clo
12	3702	55.0	4203	6 AX576404	AX576404 Sequence
13	3702	55.0	4204	6 AX576405	AX576405 Sequence
14	3702	55.0	4205	6 AX576406	AX576406 Sequence
15	3702	55.0	4216	6 AX576407	AX576407 Sequence
16	3702	55.0	4217	6 AX576408	AX576408 Sequence
17	3702	55.0	4218	6 AX576409	AX576409 Sequence
18	3384	50.3	4232	12 SYNLAICQ	L24193 Cloning vec
19	3384	50.3	4735	12 SYNBTALAC	L24367 Cloning vec

20	2729	40.6	4357	6 AR431490	AR431490 Sequence
21	2729	40.6	4357	6 AX072889	AX072889 Sequence
22	2631	40.0	4770	12 AF339022	AF339022 Expressio
23	2676	39.8	6862	6 A10361	A10361 Artificial
24	2661	39.5	6886	6 AX572872	AX572872 Sequence
25	2661	39.5	6886	6 AX573224	AX573224 Sequence
26	2658	39.5	5797	6 CQ857478	CQ857478 Sequence
27	2658	39.5	5797	6 AX786864	AX786864 Sequence
28	2648	39.4	5021	6 BD211261	BD211261 Phagocyt
29	2648	39.4	5021	6 AX008866	AX008866 Sequence
30	2648	39.4	7268	6 BD187081	BD187081 Polypepti
31	2648	39.4	7325	6 BD187078	BD187078 Polypepti
32	2648	39.4	7439	6 BD187077	BD187077 Polypepti
33	2645	39.3	5885	1 STAFRA	L42764 Staphylococ
34	2613	38.8	6806	6 AX378208	AX378208 Sequence
35	2612	38.8	4593	12 XXU02439	U02439 Cloning vec
36	2612	38.8	5075	12 U02444	U02444 Cloning vec
37	2527	37.6	4593	6 AR102988	AR102988 Sequence
38	2527	37.6	4593	6 AR172782	AR172782 Sequence
39	2527	37.6	4593	12 CVPK2332	X70478 Plasmid pKK
40	2526	37.5	5289	12 AF121783	AF121783 Expressio
41	2526	37.5	7050	12 AF121785	AF121785 Expressio
42	2491	37.0	4921	6 AX768201	AX768201 Sequence
43	2491	37.0	4935	6 AX768200	AX768200 Sequence
44	2491	37.0	4945	6 AX768202	AX768202 Sequence
45	2491	37.0	4951	6 AX768203	AX768203 Sequence

ALIGNMENTS

RESULT 1
AF050464
LOCUS AF050464 4806 bp DNA circular SYN 19-NOV-1998
DEFINITION Expression vector pKIL-HIS3, complete sequence.
ACCESSION AF050464
VERSION AF050464.1 GI:3892175
KEYWORDS
SOURCE Expression vector pKIL-HIS3
ORGANISM Expression vector pKIL-HIS3
REFERENCE 1 (bases 1 to 4806)
AUTHORS Van Reeth, T., Dreze, P.L., Szpirer, J., Szpirer, C. and Gabant, P.
TITLE Positive selection vectors to generate fused genes for the expression of his-tagged proteins
JOURNAL Biochimiques 25 (5), 898-904 (1998)
MEDLINE 99039034
REFERENCE 2 (bases 1 to 4806)
AUTHORS Van Reeth, T., Gabant, P., Dreze, P.L., Szpirer, J. and Szpirer, C.
TITLE Direct Submission
JOURNAL Submitted (24-FEB-1998) Departement de Biologie Moleculaire, Universite Libre de Bruxelles, 67, rue des Chevaux, Rhode St. Genese 1640, Belgium
FEATURES
source
1. 4806
/organism="Expression vector pKIL-HIS3"
/mol_type="genomic DNA"
/specific_host="Escherichia coli"
/db_xref="taxon:74458"
28..1002
/gene="ccdB"
28..58
/gene="ccdB"
250..774
/gene="ccdB"
/note="CcdB fusion protein"
/codon_start=1
/transl_table=11
/product="CcdB protein"
/protein_id="AAC78290.1"
/db_xref="GI:3892176"
/translation="MSRRRRQAWFSSSSSLIEGSDVLAVLQRRDWNPGVTQLRL"

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 2005, 18:49:20 ; Search time 3194 Seconds
(without alignments)
12471.490 Million cell updates/sec

Title: US-10-649-433C-1
Perfect score: 6729
Sequence: 1 gtttcacagcttatcatcga.....agttagcgcgaattgatctg 6729

Scoring table: OLIGO NUC
Gapop_60.0 , Gapext 60.0

Searched: 4390206 seqs, 2959870667 residues

Word size : 0
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3855	57.3	5027	AAx78865	Aax78865 Human tis
2	3855	57.3	5069	AAx78871	Aax78871 Human tis
3	3855	57.3	5903	ABz23939	Abz23939 Nucleotid
4	3804	56.5	4816	AAx7799	Aax7799 pTRChisB
5	3804	56.5	4816	ADf72546	Adf72546 Mouse Adi
6	3804	56.5	4816	ADg14708	Adg14708 Globular
7	3804	56.5	5013	ACA63355	Acac63355 E. coli D
8	3804	56.5	5013	AAx59423	Aax59423 E. coli pu
9	3804	56.5	5013	ADe86127	Ades86127 E. coli p
10	3804	56.5	5104	AAx7798	Aax7798 pTRChisB
11	3804	56.5	5104	ADf72545	Adf72545 Mouse Adi
12	3753	55.8	5069	AAx78874	Aax78874 Human tis
13	3753	55.8	5099	AAx78896	Aax78896 Human tis
14	3753	55.8	5132	AAx78893	Aax78893 Human tis
15	3753	55.8	5324	AAx78890	Aax78890 Human tis
16	3753	55.8	5342	AAx78893	Aax78893 Human tis
17	3753	55.8	5926	AAx32978	Aax32978 Tr7 donor
18	3753	55.8	5926	AAx32977	Aax32977 Tr7 donor
19	3753	55.8	5926	AAx45060	Aax45060 pEM delta
20	3753	55.8	5926	AAx45059	Aax45059 Transposo

C	21	3753	55.8	5926	12	ADG46817	Adg46817 Donor pla
	22	3703	55.0	4214	3	AAA59260	Aaa59260 Nucleotid
	23	3703	55.0	4214	3	AAA59260	Aaa59260 Vector pE
	24	3702	55.0	4176	1	AAx90709	Aan90709 Sequence
	25	3702	55.0	4176	12	ADI36476	Adi36476 Plasmid C
	26	3702	55.0	4203	6	ABQ74925	Abq74925 Avidity I
	27	3702	55.0	4204	6	ABQ74926	Abq74926 Avidity I
	28	3702	55.0	4205	6	ABQ74927	Abq74927 Avidity I
	29	3702	55.0	4216	6	ABQ74928	Abq74928 Avidity I
	30	3702	55.0	4217	6	ABQ74929	Abq74929 Avidity I
	31	3702	55.0	4218	6	ABQ74930	Abq74930 Avidity I
	32	3702	55.0	5926	12	ADG46818	Adg46818 pEM DNA (
	33	3683	54.7	5103	12	ADG14707	Adg14707 Mouse OBG
	34	2741	40.7	5772	10	AAx55822	Aal55822 Phi-29 po
	35	2741	40.7	5772	10	AAx55823	Aal55823 Phi-29 ex
	36	2731	40.6	4145	12	ADQ37103	Adq37103 Vector pB
	37	2731	40.6	4145	12	ADQ27625	Adq27625 pBAD/gIII
	38	2729	40.6	4357	4	AAx31389	Aaf31389 Expressio
	39	2679	39.8	4189	2	AAQ05397	Aaq05397 Secretion
	40	2661	39.5	6886	6	AAx48883	Aal48883 B subtili
	41	2661	39.5	6886	6	AAx41227	Aal41227 Pantothan
	42	2658	39.5	5797	12	ADL92656	Adl92656 Expressio
	43	2658	39.5	5797	13	ADR42002	Adr42002 VAP relat
	44	2648	39.4	5021	3	AAA07775	Aaa07775 DNA seque
	45	2648	39.4	7268	8	ABz76630	Abz76630 HIF-1 alp

ALIGNMENTS

RESULT 1
AAx78865
ID AAX78865 standard; DNA; 5027 BP.
XX
AC AAX78865;
XX
DT 08-SEP-1999 (first entry)
XX
DE Human tissue factor clone Nuv120 DNA.
XX
KW Tissue factor; human; thrombogenic; substructure; thrombose; tumour;
KW vasculative malformation; vascular endothelium; Nuv120; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 413..1171
FT /*tag= a
FT /product= "TF clone Nuv120"
XX
PN WO9932143-Al.
XX
PD 01-JUL-1999.
XX
PF 22-DEC-1998; 98WO-US027498.
XX
PR 23-DEC-1997; 97US-00996744.
XX
PA (NUVA-) NUVAS LLC.
XX
PI Houston LL, Dickinson CD;
XX
DR WPI: 1999-405116/34.
DR P-PSDB; AAY25403.
XX
PT New thrombogenic polypeptides used to, e.g. obliterate vasculative
XX malformations.
XX
PS Claim 30; Page 68-70; 97pp; English.
XX
CC This invention describes novel thrombogenic polypeptides which comprise a
CC thrombogenic substructure and a context-dependent entity which recognizes
CC desired biologically susceptible sites, e.g. tumour vascular endothelium.

CC A novel context-dependent functional entity comprises a substructure with
CC thrombogenic potential and one or more context-enhancing substructures
CC having the ability to recognize desired biologically susceptible sites,
CC where the entity imparts thrombogenic activity when positioned in the
CC function-forming-context at the biologically susceptible sites, and the
CC entity has no thrombogenic activity absent a function-forming-context at
CC the biologically susceptible sites. The context-dependent functional
CC entities impart thrombogenic activity only at biologically susceptible
CC sites. They can be used to obliterate vasculature malformations or to
CC selectively thrombose the vasculature of solid tumours. This sequence
CC encodes the human tissue factor protein Nuv120 which is used in the
XX method of the invention

SQ Sequence 5027 BP; 1287 A; 1221 C; 1321 G; 1198 T; 0 U; 0 Other;

Query Match		57.3%;	Score 3855;	DB 2;	Length 5027;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 3855;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	2875	AGCTTGGCTGTTTGGCGGATGAGAGAATTTTCAGGCTGATACAGATTAAATCAGAAC	2934		
Db	1173	AGCTTGGCTGTTTGGCGGATGAGAGAATTTTCAGGCTGATACAGATTAAATCAGAAC	1232		
QY	2935	GCAGAAGCGGCTGTATAAACAGAAATTTGCCCTGGCGGAGTAGCGCGTGGTCCACCTG	2994		
Db	1233	GCAGAAGCGGCTGTATAAACAGAAATTTGCCCTGGCGGAGTAGCGCGTGGTCCACCTG	1292		
QY	2995	ACCCCATGCCAACTCAGAAAGTGAACCGCGTAGCGCGATGGTAGTGTGGGCTCTCCCC	3054		
Db	1293	ACCCCATGCCAACTCAGAAAGTGAACCGCGTAGCGCGATGGTAGTGTGGGCTCTCCCC	1352		
QY	3055	ATGCAGAGTAGGGAACCTGCCAGGCATCAAAATAAAACGAAAGGCTCAGTCGAAAGACTGG	3114		
Db	1353	ATGCAGAGTAGGGAACCTGCCAGGCATCAAAATAAAACGAAAGGCTCAGTCGAAAGACTGG	1412		
QY	3115	GCCTTTCTGTTTATCTGTTGTTGTCGTTGACGCTCTCCTGAGTAGGACAAATCCGCGG	3174		
Db	1413	GCCTTTCTGTTTATCTGTTGTTGTCGTTGACGCTCTCCTGAGTAGGACAAATCCGCGG	1472		
QY	3175	GGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGCGAGGAGCGCCGCCA	3234		
Db	1473	GGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGCGAGGAGCGCCGCCA	1532		
QY	3235	TAACTGCCAGGCATCAAAATTAAGCAGAAAGGCGCATCTGACGGATGGCCTTTTTCGCTTT	3294		
Db	1533	TAACTGCCAGGCATCAAAATTAAGCAGAAAGGCGCATCTGACGGATGGCCTTTTTCGCTTT	1592		
QY	3295	CTACAAACTCTTTTGTGTTTATTTTCTAAATACATTCAAATATGATCCGCTCATGAGAC	3354		
Db	1593	CTACAAACTCTTTTGTGTTTATTTTCTAAATACATTCAAATATGATCCGCTCATGAGAC	1652		
QY	3355	AATAACCTGTAAATGCTTCAATAATATTTGAAAAGGAAGATGATGATTTCAACATT	3414		
Db	1653	AATAACCTGTAAATGCTTCAATAATATTTGAAAAGGAAGATGATGATTTCAACATT	1712		
QY	3415	TCCGTTGCGCCCTTATTCCTTTTGGGCAATTTTGGCCTCTCTGTTTGTCTACCCAG	3474		
Db	1713	TCCGTTGCGCCCTTATTCCTTTTGGGCAATTTTGGCCTCTCTGTTTGTCTACCCAG	1772		
QY	3475	AAACCGCTGTGAAAGTAAAGATGCTGAAGATCAGTTGGTGCACAGTGGGTACATCG	3534		
Db	1773	AAACCGCTGTGAAAGTAAAGATGCTGAAGATCAGTTGGTGCACAGTGGGTACATCG	1832		
QY	3535	AACTGGATCTAACAGCGGTAAAGTCTTGTAGAGTTTTTGCCTCCCGGAGAAACGTTTTCCAA	3594		
Db	1833	AACTGGATCTAACAGCGGTAAAGTCTTGTAGAGTTTTTGCCTCCCGGAGAAACGTTTTCCAA	1892		
QY	3595	TGATGAGCACTTTTAAAGTCTGCTATGTGGCGCGGTATTATCCCGTGTGACGCGGGC	3654		
Db	1893	TGATGAGCACTTTTAAAGTCTGCTATGTGGCGCGGTATTATCCCGTGTGACGCGGGC	1952		
QY	3655	AAGAGCAACTCCGTCGCGCATACACTATTCTCAGAAATGACTTGGTTGAGTACTCACCAAG	3714		

Db	1953	AAGACCAACTCGGTGCGCGCATACACTATTCTCAGAAATGACTTGGTTAGTACTCACCAAG	2012		
QY	3715	TCACAGAAAAGCATCTTTACGGATGGCATGACAGTAAGAGAATTAATGACAGTGTGTCATAA	3774		
Db	2013	TCACAGAAAAGCATCTTTACGGATGGCATGACAGTAAGAGAATTAATGACAGTGTGTCATAA	2072		
QY	3775	CCATGAGTGATAACACTGCGGCGCAACTTACTCTCTGACAAAGATCGGAGGACCGAAGGAGC	3834		
Db	2073	CCATGAGTGATAACACTGCGGCGCAACTTACTCTCTGACAAAGATCGGAGGACCGAAGGAGC	2132		
QY	3835	TAAACCGCTTTTTCACAAACATGCGGGATCATGTAACCTCGCTTGATCGTTGGGAACCGG	3894		
Db	2133	TAAACCGCTTTTTCACAAACATGCGGGATCATGTAACCTCGCTTGATCGTTGGGAACCGG	2192		
QY	3895	AGCTGAATGAAGCCATACCAACGACGAGCGTGACACCAAGATGCTGTAGCAATGGCAA	3954		
Db	2193	AGCTGAATGAAGCCATACCAACGACGAGCGTGACACCAAGATGCTGTAGCAATGGCAA	2252		
QY	3955	CAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTA	4014		
Db	2253	CAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTA	2312		
QY	4015	TAGACTGATGAGCGCGGATAAAGTTGACGACCACTTCTGCGCTCGGCGCTTCCGGCTG	4074		
Db	2313	TAGACTGATGAGCGCGGATAAAGTTGACGACCACTTCTGCGCTCGGCGCTTCCGGCTG	2372		
QY	4075	GCTGGTTTATTTGCTGATAAAATCTGGAGCCGAGCGGTGCTCGCGGTATCATTTGAG	4134		
Db	2373	GCTGGTTTATTTGCTGATAAAATCTGGAGCCGAGCGGTGCTCGCGGTATCATTTGAG	2432		
QY	4135	CACCTGGGCGGATGTTAAGCCCTCCGCTATGCTAGTTATCTACACGACGGGAGTCAAG	4194		
Db	2433	CACCTGGGCGGATGTTAAGCCCTCCGCTATGCTAGTTATCTACACGACGGGAGTCAAG	2492		
QY	4195	CAACTATGATGAACAAATAGACAGATCGCTGAGATAGGTGCTCTACTGATTAAGCATTT	4254		
Db	2493	CAACTATGATGAACAAATAGACAGATCGCTGAGATAGGTGCTCTACTGATTAAGCATTT	2552		
QY	4255	GGTAACTGTGACGCAAGTTTATCTCATATATATCTTTAGATTGATTTTAAATCTTCAATTTT	4314		
Db	2553	GGTAACTGTGACGCAAGTTTATCTCATATATATCTTTAGATTGATTTTAAATCTTCAATTTT	2612		
QY	4315	AATTTAAAGGATCTTAGTGGAAGATCCCTTTTGTGATAATCTCATGACCAAAATCCCTTAAC	4374		
Db	2613	AATTTAAAGGATCTTAGTGGAAGATCCCTTTTGTGATAATCTCATGACCAAAATCCCTTAAC	2672		
QY	4375	GTEAGTTTTTCTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAGGATCTTCTTGAG	4434		
Db	2673	GTGAGTTTTTCTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAGGATCTTCTTGAG	2732		
QY	4435	ATCCTTTTTTCTGCGCGTAAATCTGCTGCTTGCAAAACAAAAACACCGCTACCGCGG	4494		
Db	2733	ATCCTTTTTTCTGCGCGTAAATCTGCTGCTTGCAAAACAAAAACACCGCTACCGCGG	2792		
QY	4495	TGGTTGTTTTCGCGGATCAAGAGCTACCAACTCTTTTTCGAAAGGTAACTGGCTTCAGCA	4554		
Db	2793	TGGTTGTTTTCGCGGATCAAGAGCTACCAACTCTTTTTCGAAAGGTAACTGGCTTCAGCA	2852		
QY	4555	GAGCGAGATACCAAAATCTGCTCTCTAGTGTAGCCGCTAGTTAGGCCACCACTTCAAGA	4614		
Db	2853	GAGCGAGATACCAAAATCTGCTCTCTAGTGTAGCCGCTAGTTAGGCCACCACTTCAAGA	2912		
QY	4615	ACTCTGTAGCACCGCTACATACCTCGCTCTGCTTAATCTGTTACAGTGGCTGTGCCA	4674		
Db	2913	ACTCTGTAGCACCGCTACATACCTCGCTCTGCTTAATCTGTTACAGTGGCTGTGCCA	2972		
QY	4675	GTGGCGATAAGTCTGCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGC	4734		
Db	2973	GTGGCGATAAGTCTGCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGC	3032		
QY	4735	AGCGGTGCGGCTGAACGCGGGGTTCTGTGCACACAGCCAGCTTGAGCGAAACGACTACA	4794		
Db	3033	AGCGGTGCGGCTGAACGCGGGGTTCTGTGCACACAGCCAGCTTGAGCGAAACGACTACA	3092		

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 2005, 12:59:49 ; Search time 992 Seconds
(without alignments)
11099.300 Million cell updates/sec

Title: US-10-649-433C-1

Perfect score: 6729

Sequence: 1 gtttcacagcttatcatcga.....agttacgcgaattgatcgtg 6729

Scoring table: OLIGO NUC

Gapop_60.0 , Gapext 60.0

Searched: 1202784 seqs, 818138359 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfileseq1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3804	56.5	5013	US-09-183-188B-5	Sequence 5, Appli
2	3753	55.8	5926	US-09-027-169-3	Sequence 3, Appli
3	3753	55.8	5926	US-09-027-169-4	Sequence 4, Appli
4	3752	55.8	4476	US-08-801-344-2	Sequence 2, Appli
5	3752	55.8	4476	US-09-498-599-2	Sequence 2, Appli
6	2729	40.6	4357	US-09-795-872-10	Sequence 10, Appli
7	2527	37.6	4593	US-08-801-344-1	Sequence 1, Appli
8	2527	37.6	4593	US-09-498-599-1	Sequence 1, Appli
9	2433	36.2	5238	5453363-1	Patent No. 5453363
10	2433	36.2	5238	6 5453363-1	Sequence 3, Appli
11	2338	34.7	6312	US-08-531-601-3	Sequence 3, Appli
12	2338	34.7	6312	US-08-859-032-3	Sequence 3, Appli
13	2329	34.6	5446	US-09-358-856C-1	Sequence 1, Appli
14	2179	32.4	9144	US-08-556-978B-79	Sequence 79, Appli
15	2130	31.7	7731	US-09-301-593-29	Sequence 29, Appli
16	2130	31.7	7731	US-09-301-593-42	Sequence 42, Appli
17	2130	31.7	7892	US-07-916-098A-40	Sequence 40, Appli
18	2129	31.6	3699	US-08-053-131-120	Sequence 120, App
19	2129	31.6	3699	US-08-845-641-120	Sequence 120, App
20	2129	31.6	3699	US-07-853-408B-120	Sequence 120, App
21	2129	31.6	3699	US-08-096-762-120	Sequence 120, App
22	2129	31.6	3699	US-08-308-865-120	Sequence 120, App
23	2129	31.6	3699	PCT-US92-10983-120	Sequence 120, App
24	2129	31.6	5115	US-08-825-852-19	Sequence 19, Appli
25	2129	31.6	5115	US-09-052-888-19	Sequence 19, Appli
26	2129	31.6	5115	US-09-723-890-19	Sequence 19, Appli
27	2129	31.6	5115	US-09-723-901-19	Sequence 19, Appli

ALIGNMENTS

RESULT 1

US-09-183-188B-5
; Sequence 5, Application US/09183188B
; Patent No. 6491905

GENERAL INFORMATION:
; APPLICANT: Sorscher, Eric J.

; APPLICANT: Parker, William B.

; APPLICANT: Waud, William

; APPLICANT: Gadi, Vijaykrishna K.

; TITLE OF INVENTION: RECOMBINANT E. COLI FOR DELIVERY OF PNP TO TUMOR CELLS

; FILE REFERENCE: UAB-12405/22

; CURRENT APPLICATION NUMBER: US/09/183,188B

; PRIOR FILING DATE: 1998-10-30

; PRIOR APPLICATION NUMBER: 08/702,181

; PRIOR FILING DATE: 1996-08-23

; PRIOR APPLICATION NUMBER: 08/122,321

; PRIOR FILING DATE: 1993-09-14

; PRIOR APPLICATION NUMBER: 60/064,676

; PRIOR FILING DATE: 1997-10-31

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 5

; LENGTH: 5013

; TYPE: DNA

; ORGANISM: Escherichia coli

US-09-183-188B-5

Query Match 56.5%; Score 3804; DB 4; Length 5013;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 3854; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	2875	AGCTTGGCTGTTTGGCGGATGAGAGAGATTTCAGCCTGATACAGATTAAATCAGAAC	2934
Db	1159	AGCTTGGCTGTTTGGCGGATGAGAGAGATTTCAGCCTGATACAGATTAAATCAGAAC	1218
QY	2935	GCAGAACGGCTGTGATATAAACAGAAATTCCTGGCGGAGTAGCGCGGTGTCACCTG	2994
Db	1219	GCAGAACGGCTGTGATATAAACAGAAATTCCTGGCGGAGTAGCGCGGTGTCACCTG	1278
QY	2995	ACCCATGCCGAACCTCAGAAAGTGAACCGCGGTAGCGCGGTGTCAGTGTGGGTCTCCCC	3054
Db	1279	ACCCATGCCGAACCTCAGAAAGTGAACCGCGGTAGCGCGGTGTCAGTGTGGGTCTCCCC	1338
QY	3055	ATGCCAGAGTAGGGAACCTCCAGGATCAATATAAACGAAAGGCTCAGTCGAAAGACTGG	3114
Db	1339	ATGCCAGAGTAGGGAACCTCCAGGATCAATATAAACGAAAGGCTCAGTCGAAAGACTGG	1398
QY	3115	GCCTTTCGTTTTATCTGTTGTTTGTGGTGAACGCTCTCTGAGTAGGACAAATCGCGCG	3174
Db	1399	GCCTTTCGTTTTATCTGTTGTTTGTGGTGAACGCTCTCTGAGTAGGACAAATCGCGCG	1458

Qy	3175	GGAGCGGATTTGAACGTTTGGCAAGCAACGCGCCGGAGGGTGGCGGCGAGGACGCCCGCCCA	3234
Db	1459	GGAGCGGATTTGAACGTTTGGCAAGCAACGCGCCGGAGGGTGGCGGCGAGGACGCCCGCCCA	1518
Qy	3235	TAAACTCCAGGCAATCAAAATTAAGCAGAAAGCCATCCTGACGATGGCCCTTTTTCGGTTT	3294
Db	1519	TAAACTCCAGGCAATCAAAATTAAGCAGAAAGCCATCCTGACGATGGCCCTTTTTCGGTTT	1578
Qy	3295	CTACAAACTCTTTTGTGTTATTTTCTTAAATACAAATATGTATCGCTCATGAGAC	3354
Db	1579	CTACAAACTCTTTTGTGTTATTTTCTTAAATACAAATATGTATCGCTCATGAGAC	1638
Qy	3355	AATACCCCTGATAAAATGCTTCAATAATTTGAAAAAGGAGATGATGAGTATTTCAACATT	3414
Db	1639	AATACCCCTGATAAAATGCTTCAATAATTTGAAAAAGGAGATGATGAGTATTTCAACATT	1698
Qy	3415	TCGGTGTGGCCCTTATTTCCCTTTTTCGGGCATTTTGGCCCTTCTCTGTTTTTGTCAACCAG	3474
Db	1699	TCGGTGTGGCCCTTATTTCCCTTTTTCGGGCATTTTGGCCCTTCTCTGTTTTTGTCAACCAG	1758
Qy	3475	AAACGCTGGTGAAGATGCTGTAAGATCAGTTTGGGTGCAAGTGGGTTCATCG	3534
Db	1759	AAACGCTGGTGAAGATGCTGTAAGATCAGTTTGGGTGCAAGTGGGTTCATCG	1818
Qy	3535	AACCTGGATCTCAACAGCGGTAAAGATCCTTCAGAGTTTTTGCSCCGAAGAACGTTTTCCAA	3594
Db	1819	AACCTGGATCTCAACAGCGGTAAAGATCCTTCAGAGTTTTTGCSCCGAAGAACGTTTTCCAA	1878
Qy	3595	TGATGAGCACTTTTAAAGTTCTGCTATGTGGCGGTTATATCCCGTGTGACCGCGGGC	3654
Db	1879	TGATGAGCACTTTTAAAGTTCTGCTATGTGGCGGTTATATCCCGTGTGACCGCGGGC	1938
Qy	3655	AAGAGCAACTCGGTGCGCGCATACACTATTTCTCAGATGACTTGGTTGAGTACTACCCAG	3714
Db	1939	AAGAGCAACTCGGTGCGCGCATACACTATTTCTCAGATGACTTGGTTGAGTACTACCCAG	1998
Qy	3715	TCAAGAAAAAGCATCTTTACGGATGGCATGACAGTAAGAGAAATATGACAGTGTGCCATAA	3774
Db	1999	TCAAGAAAAAGCATCTTTACGGATGGCATGACAGTAAGAGAAATATGACAGTGTGCCATAA	2058
Qy	3775	CCATGAGTGATAACACTGCGGCCAACTTACTTGACAAAGATCGAGGACCGAAGGAGC	3834
Db	2059	CCATGAGTGATAACACTGCGGCCAACTTACTTGACAAAGATCGAGGACCGAAGGAGC	2118
Qy	3835	TAAACCGCTTTTTCGACAAATGGGGATCATGTAACTCGCCCTTGATCGTTTGGAAACCGG	3894
Db	2119	TAAACCGCTTTTTCGACAAATGGGGATCATGTAACTCGCCCTTGATCGTTTGGAAACCGG	2178
Qy	3895	AGCTGAATGAAGCCATACCAAAAGCAGCGGTGACACCCAGATGCCTGTAGCAATGGCAA	3954
Db	2179	AGCTGAATGAAGCCATACCAAAAGCAGCGGTGACACCCAGATGCCTGTAGCAATGGCAA	2238
Qy	3955	CAACGTTTGGCAAACTAATTAACCTGGCAACTACTTACTTACTAGCTTCCCGGCAACAATTA	4014
Db	2239	CAACGTTTGGCAAACTAATTAACCTGGCAACTACTTACTTACTAGCTTCCCGGCAACAATTA	2298
Qy	4015	TAGA CTGGATGGAGCGGATAAAGTTTCGAGGACCACTTCTCGCGCTCGGCCCTTTCGGCTG	4074
Db	2299	TAGA CTGGATGGAGCGGATAAAGTTTCGAGGACCACTTCTCGCGCTCGGCCCTTTCGGCTG	2358
Qy	4075	GCTGGTTTTATGCTGATAAATCTGGAGCCGCTGAGCGGTCTCGCGGTATCATTTGCAG	4134
Db	2359	GCTGGTTTTATGCTGATAAATCTGGAGCCGCTGAGCGGTCTCGCGGTATCATTTGCAG	2418
Qy	4135	CACCTGGGGCCAGATGGTAAGCCCTCCCGTATCTGTAGTTATCTACAGACGGGAGTCAGG	4194
Db	2419	CACCTGGGGCCAGATGGTAAGCCCTCCCGTATCTGTAGTTATCTACAGACGGGAGTCAGG	2478
Qy	4195	CAACTATGGATGAACGAAAAATAGACAGATCGCTGAGATAGTGTGCTCTACTGATTAAGCATT	4254
Db	2479	CAACTATGGATGAACGAAAAATAGACAGATCGCTGAGATAGTGTGCTCTACTGATTAAGCATT	2538

Qy	4255	GGTAACTGTGACAGCCAAAGTTTACTCATATATACATTGATTTGATTTAAACCTTCATATTTT	4314
Db	2539	GGTAACTGTGACAGCCAAAGTTTACTCATATATACATTGATTTGATTTAAACCTTCATATTTT	2598
Qy	4315	AATTTAAAAGGATCTAGGTGAAGATCCCTTTTGTAAATCTCATGACCAAAATCCCTTAAC	4374
Db	2599	AATTTAAAAGGATCTAGGTGAAGATCCCTTTTGTAAATCTCATGACCAAAATCCCTTAAC	2658
Qy	4375	GTGAGTTTTCGTTCCACTGAGCGGTGAGACCCCGTAGAAAGAATCAAAGATCTTCTTGAG	4434
Db	2659	GTGAGTTTTCGTTCCACTGAGCGGTGAGACCCCGTAGAAAGAATCAAAGATCTTCTTGAG	2718
Qy	4435	ATCCTTTTTCCTGCGCGTAATCTGCTGCTGCTCAACCAAAAAACACCGCTACACAGCGG	4494
Db	2719	ATCCTTTTTCCTGCGCGTAATCTGCTGCTGCTCAACCAAAAAACACCGCTACACAGCGG	2778
Qy	4495	TGTTTGTGTTGCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACTGGCTTCAGCA	4554
Db	2779	TGTTTGTGTTGCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACTGGCTTCAGCA	2833
Qy	4555	GAGCGCAGATACCAATATCTGTCCTTCTAGTGTAGCGGTAGTTAGGCCACACACTTCAGA	4614
Db	2839	GAGCGCAGATACCAATATCTGTCCTTCTAGTGTAGCGGTAGTTAGGCCACACACTTCAGA	2898
Qy	4615	ACTCTGTAGCACCGCTCATACACTCGCTCTGCTAATCCTGTTTACAGTGGCTGCTGCCA	4674
Db	2899	ACTCTGTAGCACCGCTCATACACTCGCTCTGCTAATCCTGTTTACAGTGGCTGCTGCCA	2958
Qy	4675	GTGGCGATTAAGTGTGTCTTACCGGGTTGGACTCAAGACGATAGTTTACGGATTAAGGCGC	4734
Db	2959	GTGGCGATTAAGTGTGTCTTACCGGGTTGGACTCAAGACGATAGTTTACGGATTAAGGCGC	3018
Qy	4735	AGCGGTGCGGCTGAAACGGGGGGTTCGTGCACACAGCCAGCTTGTGAGCGAACGACTACA	4794
Db	3019	AGCGGTGCGGCTGAAACGGGGGGTTCGTGCACACAGCCAGCTTGTGAGCGAACGACTACA	3078
Qy	4795	CCGAACTGAGATACCTPACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGAGAA	4854
Db	3079	CCGAACTGAGATACCTPACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGAGAA	3138
Qy	4855	AGCGCAGCAGGTATCCGTTAAGCGCCGGGTGCGAACAGGAGAGCGCACGAGGAGCTTC	4914
Db	3139	AGCGCAGCAGGTATCCGTTAAGCGCCGGGTGCGAACAGGAGAGCGCACGAGGAGCTTC	3198
Qy	4915	CAGGGGAAACGCGCTGGTATCTTTATAGTCTGTGCGGGTTTCGCCACCTCTGACTGAGC	4974
Db	3199	CAGGGGAAACGCGCTGGTATCTTTATAGTCTGTGCGGGTTTCGCCACCTCTGACTGAGC	3258
Qy	4975	GTCCGATTTTGTGATGCTCGTACAGGGGGCGAGCCTATGGAATAAACGCCAGCAACGCGG	5034
Db	3259	GTCCGATTTTGTGATGCTCGTACAGGGGGCGAGCCTATGGAATAAACGCCAGCAACGCGG	3318
Qy	5035	CCTTTTTACGGTTCCTGGGCTTTTGTGGCCCTTTTGTCTCACATGTTCTTTCTCGGCTTAT	5094
Db	3319	CCTTTTTACGGTTCCTGGGCTTTTGTGGCCCTTTTGTCTCACATGTTCTTTCTCGGCTTAT	3378
Qy	5095	CCCTGATTCGTGAGATAACCGTATPACCGCTTTGTAGTGAAGTGAATACCGCTCGCGCA	5154
Db	3379	CCCTGATTCGTGAGATAACCGTATPACCGCTTTGTAGTGAAGTGAATACCGCTCGCGCA	3438
Qy	5155	GCGCAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAACGGAAGCGCCCTGATGCGGT	5214
Db	3439	GCGCAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAACGGAAGCGCCCTGATGCGGT	3498
Qy	5215	ATTTTCTCCTTACGCAATCTGTGCGGTATTTTCAACCGCATATGGTGCACTCTCAGTACAA	5274
Db	3499	ATTTTCTCCTTACGCAATCTGTGCGGTATTTTCAACCGCATATGGTGCACTCTCAGTACAA	3558
Qy	5275	TCGTCTGTAGTCGCGCATATGTTAAGCCAGTATACACTCCGCTATCGCTACGTTAGCTGGT	5334
Db	3559	TCGTCTGTAGTCGCGCATATGTTAAGCCAGTATACACTCCGCTATCGCTACGTTAGCTGGT	3618
Qy	5335	CATGGCTGGCGCCCGGACACCCGCGCAACACCCGCTGACCGCGCCCTGACGGGCTTGTCTGT	5394

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 2005, 18:42:20 ; Search time 3776 Seconds
(without alignments)
11551.793 Million cell updates/sec

Title: US-10-649-433C-1
Perfect score: 6729
Sequence: 1 gtttcacagcttatcatcg.....agtttagcggaattgatctg 6729

Scoring table: OLIGO NUC
Gapop_60.0 , Gapext 60.0

Searched: 7297361 seqs, 3241162794 residues

Word size : 0
Total number of hits satisfying chosen parameters: 14594722

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*

3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*

7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*

10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*

11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*

12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*

13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*

14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*

15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*

16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*

17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*

18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*

19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:*

20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:*

21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:*

22: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*

23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq:*

24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*

25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*

26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6729	100.0	6729	21	US-10-649-433C-1
2	3804	56.5	5013	14	Sequence 1, Appli
3	3804	56.5	5013	15	Sequence 5, Appli
C 4	3753	55.8	5926	13	Sequence 5, Appli
C 5	3753	55.8	5926	13	Sequence 3, Appli
6	3702	55.0	4176	16	Sequence 4, Appli
7	2741	40.7	5772	20	Sequence 3, Appli
					Sequence 37, Appli

8	2741	40.7	5772	20	US-10-131-998A-38	Sequence 38, Appli
9	2731	40.6	4145	20	US-10-699-088-163	Sequence 163, App
10	2731	40.6	4145	21	US-10-699-114-163	Sequence 163, App
11	2729	40.6	4357	10	US-09-795-872-10	Sequence 10, Appl
12	2729	40.6	4357	19	US-10-662-824-10	Sequence 10, Appl
13	2661	39.5	6886	18	US-10-466-642-20	Sequence 20, Appl
14	2661	39.5	6886	18	US-10-466-717-20	Sequence 20, Appl
15	2658	39.5	5797	17	US-10-316-194-167	Sequence 167, App
16	2658	39.5	5797	21	US-10-732-180-203	Sequence 203, App
17	2613	38.8	6806	10	US-09-908-943A-194	Sequence 194, App
18	2613	38.8	6806	20	US-10-801-487-194	Sequence 194, App
19	2613	38.8	6806	20	US-10-801-938-194	Sequence 194, App
20	2613	38.8	6806	20	US-10-801-509-194	Sequence 194, App
21	2613	38.8	6806	20	US-10-801-486-194	Sequence 194, App
22	2613	38.8	6806	21	US-10-801-493-194	Sequence 194, App
23	2491	37.0	4921	21	US-10-495-491-2	Sequence 2, Appli
24	2491	37.0	4935	21	US-10-495-491-1	Sequence 1, Appli
25	2491	37.0	4945	21	US-10-495-491-3	Sequence 3, Appli
26	2491	37.0	4951	21	US-10-495-491-4	Sequence 4, Appli
27	2484	36.9	5398	22	US-10-505-082-3	Sequence 3, Appli
28	2329	34.6	5446	17	US-10-393-269-1	Sequence 1, Appli
29	2279	33.9	6464	15	US-10-151-690-20	Sequence 20, Appli
C 30	2260	33.6	7615	10	US-09-898-616A-9	Sequence 9, Appli
C 31	2260	33.6	7615	17	US-10-187-498A-9	Sequence 9, Appli
32	2227	33.1	5522	10	US-09-872-868-16	Sequence 16, Appli
33	2227	33.1	5525	10	US-09-872-868-17	Sequence 17, Appli
34	2227	33.1	6086	10	US-09-872-868-12	Sequence 12, Appli
35	2227	33.1	6086	10	US-09-872-868-13	Sequence 13, Appli
36	2227	33.1	6086	10	US-09-872-868-14	Sequence 14, Appli
37	2227	33.1	6088	10	US-09-872-868-15	Sequence 15, Appli
38	2130	31.7	7731	9	US-09-301-593-29	Sequence 29, Appli
39	2130	31.7	7731	9	US-09-301-593-42	Sequence 42, Appli
40	2130	31.7	7731	15	US-10-159-006-29	Sequence 29, Appli
41	2130	31.7	7731	15	US-10-159-006-42	Sequence 42, Appli
C 42	2130	31.7	9077	9	US-09-734-300-1	Sequence 1, Appli
C 43	2130	31.7	9077	9	US-09-734-300-3	Sequence 1, Appli
C 44	2129	31.6	4010	16	US-10-353-908-1	Sequence 1, Appli
C 45	2129	31.6	5024	17	US-10-313-963A-52	Sequence 52, Appli

ALIGNMENTS

RESULT 1
US-10-649-433C-1
; Sequence 1, Application US/10649433C
; Publication No. US20050118726A1
; GENERAL INFORMATION:
; APPLICANT: Schultz, Jerome S.
; TITLE OF INVENTION: System and Method for Detecting Bioanalytes and Method for
; TITLE OF INVENTION: Producing a Bioanalyte Sensor
; FILE REFERENCE: 03-016
; CURRENT APPLICATION NUMBER: US/10/649,433C
; CURRENT FILING DATE: 2003-08-26
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 6729
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-649-433C-1

Query Match	100.0%	Score 6729;	DB 21;	Length 6729;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 6729;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	GTTCACAGCTTATCATCGACGTGACCGTGACCAATGCTTCTGGCGTCAGCAGCCATC	60	
Db	1	GTTCACAGCTTATCATCGACGTGACCGTGACCAATGCTTCTGGCGTCAGCAGCCATC	60	
QY	61	GGAAGCTGTGGTATGGCTGTGTCAGGTGCGTAATCACTGCAATATTCGTTCGCTCAAGGC	120	
Db	61	GGAAGCTGTGGTATGGCTGTGTCAGGTGCGTAATCACTGCAATATTCGTTCGCTCAAGGC	120	

QY 121 GCATCCCGTTCCTGGATAATGTTTTTTGGCCGACATCATACGGTTCCTGGCAATATTC 180
DB 121 GCATCCCGTTCCTGGATAATGTTTTTTGGCCGACATCATACGGTTCCTGGCAATATTC 180
QY 181 TGAAATGAGCTGTTGACAAATTAATCATCCGGCTCGTATAATGTGTGGAATGTGAGCGGA 240
DB 181 TGAAATGAGCTGTTGACAAATTAATCATCCGGCTCGTATAATGTGTGGAATGTGAGCGGA 240
QY 241 TAACAATTTACACAGGAACAGCCGCTGAGAAAAAGCGAAGCGGCATCTGCTCTTTAA 300
DB 241 TAACAATTTACACAGGAACAGCCGCTGAGAAAAAGCGAAGCGGCATCTGCTCTTTAA 300
QY 301 CAATTTATCAGCAATCTGTGTGGCCTCGACCCGGAATATCGAATTAACCTTTATATTA 360
DB 301 CAATTTATCAGCAATCTGTGTGGCCTCGACCCGGAATATCGAATTAACCTTTATATTA 360
QY 361 AAAATTAAGAGGTATATATTAATGTATCGAATTAAGAGGAATAAACCATGGTGAG 420
DB 361 AAAATTAAGAGGTATATATTAATGTATCGAATTAAGAGGAATAAACCATGGTGAG 420
QY 421 CAAAGGCGAGGAGCTGTTACCGGGGTGGTCCCATCTCTGTGTGAGCTGGACGGCGA 480
DB 421 CAAAGGCGAGGAGCTGTTACCGGGGTGGTCCCATCTCTGTGTGAGCTGGACGGCGA 480
QY 481 AAACGGCCACAAGTTTCAGGGTGTCCGGGAGGGCGAGGCGATGCCATACGGCAAGCT 540
DB 481 AAACGGCCACAAGTTTCAGGGTGTCCGGGAGGGCGAGGCGATGCCATACGGCAAGCT 540
QY 541 GACCTGGAAGTTTCATCTGCACACCGCAAGCTGCCGCTGGCCCGCCACCGCTCGTGAC 600
DB 541 GACCTGGAAGTTTCATCTGCACACCGCAAGCTGCCGCTGGCCCGCCACCGCTCGTGAC 600
QY 601 CACCTTCGGCTAGGCGCTGCAGTGTTCGCCGCTACCCGACACACATGAGACAGACA 660
DB 601 CACCTTCGGCTAGGCGCTGCAGTGTTCGCCGCTACCCGACACACATGAGACAGACA 660
QY 661 CTTCTTCAAGTCCGCGATCCCGAAGGCTACGTCCAGGAGCGACCATCTCTTCAAGGA 720
DB 661 CTTCTTCAAGTCCGCGATCCCGAAGGCTACGTCCAGGAGCGACCATCTCTTCAAGGA 720
QY 721 CGACGGCAACTACAAGACCGCGCGAGGTGAAGTTTCGAGGGCGACACCTGGTGAA 780
DB 721 CGACGGCAACTACAAGACCGCGCGAGGTGAAGTTTCGAGGGCGACACCTGGTGAA 780
QY 781 CATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCTCTGGGCAACAAGCTGA 840
DB 781 CATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCTCTGGGCAACAAGCTGA 840
QY 841 GTACAACTACAAGCGCAACAGTCTATATCATGGCCGACAGCAGAGAACGGCATCA 900
DB 841 GTACAACTACAAGCGCAACAGTCTATATCATGGCCGACAGCAGAGAACGGCATCA 900
QY 901 GGTGAATTTCAAGATCCGCGCAACATCGAGGACGGCAGCTGCGAGCTGCGCGACCTA 960
DB 901 GGTGAATTTCAAGATCCGCGCAACATCGAGGACGGCAGCTGCGAGCTGCGCGACCTA 960
QY 961 CCAGCAGAACACCCCATCGGCGACGGCCCGTGTCTGCTGCTCCGACCAACCACTACCTGAG 1020
DB 961 CCAGCAGAACACCCCATCGGCGACGGCCCGTGTCTGCTGCTCCGACCAACCACTACCTGAG 1020
QY 1021 CTAACGCTCGGCTGAGCAAAAGACCCCAACAGAGAGCGCGATCATAGTGTCTGCTGGA 1080
DB 1021 CTAACGCTCGGCTGAGCAAAAGACCCCAACAGAGAGCGCGATCATAGTGTCTGCTGGA 1080
QY 1081 GTTCGTGACCGCGCGGATCACTCTCGGATGGACAGCTGTACAAGACTAGTGTGA 1140
DB 1081 GTTCGTGACCGCGCGGATCACTCTCGGATGGACAGCTGTACAAGACTAGTGTGA 1140
QY 1141 TACTCGCATTTGGTGAACAATCTATAAGTACGACGATATCTTGTGTAGTGGCA 1200
DB 1141 TACTCGCATTTGGTGAACAATCTATAAGTACGACGATATCTTGTGTAGTGGCA 1200

QY 1201 GGCTATTGAGCAAGATCGAAAGCCGCGCAGATGTTTCAGCTGCTGATGAATGATTTCTCA 1260
DB 1201 GGCTATTGAGCAAGATCGAAAGCCGCGCAGATGTTTCAGCTGCTGATGAATGATTTCTCA 1260
QY 1261 GAATGACAGCTCCAAAGCAGAACGATCAGATCGATTTGCTGGCCAAAGGGGTGAAGGC 1320
DB 1261 GAATGACAGCTCCAAAGCAGAACGATCAGATCGATTTGCTGGCCAAAGGGGTGAAGGC 1320
QY 1321 ACTGGCCATCAACCTGGTGTGACCCGCGCAGCTGCGGGTACCGGTGATTGAGAAAGCGCGTG 1380
DB 1321 ACTGGCCATCAACCTGGTGTGACCCGCGCAGCTGCGGGTACCGGTGATTGAGAAAGCGCGTG 1380
QY 1381 GCAAAAAGCTGCGGTGTTTTTTTCAACAAAGAAACCGTCTCGTAAGGCGCTGGAATAGCTA 1440
DB 1381 GCAAAAAGCTGCGGTGTTTTTTTCAACAAAGAAACCGTCTCGTAAGGCGCTGGAATAGCTA 1440
QY 1441 CGACAAAGCCTTACTAGTTGGCCTGACTCAAAAGAGTCCGGCATTTATCAAGGCGATTT 1500
DB 1441 CGACAAAGCCTTACTAGTTGGCCTGACTCAAAAGAGTCCGGCATTTATCAAGGCGATTT 1500
QY 1501 GATTGCTAAACACTGGGCGCGCAATCAGGGTTGGGATCTGAACAAAGACGCTCAGATTTCA 1560
DB 1501 GATTGCTAAACACTGGGCGCGCAATCAGGGTTGGGATCTGAACAAAGACGCTCAGATTTCA 1560
QY 1561 GTTGTACTGCTGAAAAGGTGAACCGGGCATCCCGATGCAAGACACGCTACACCTTACGT 1620
DB 1561 GTTGTACTGCTGAAAAGGTGAACCGGGCATCCCGATGCAAGACACGCTACACCTTACGT 1620
QY 1621 GATTAAAGAAATTGAACGATAAAGGCATCAAAACCTGAAACAGTTACAGTTAGATACCGCAAT 1680
DB 1621 GATTAAAGAAATTGAACGATAAAGGCATCAAAACCTGAAACAGTTACAGTTAGATACCGCAAT 1680
QY 1681 GTGGGACACCGCTCAGGCGAAAGATAAGATGGAAGCTGCTGGCTGTCTGCGCGAAGCCAA 1740
DB 1681 GTGGGACACCGCTCAGGCGAAAGATAAGATGGAAGCTGCTGGCTGTCTGCGCGAAGCCAA 1740
QY 1741 CAAATCGAAGTGGTTATCGCCAAACAGATGCGATGGCAATGGCGCGGTTGAAGCGCT 1800
DB 1741 CAAATCGAAGTGGTTATCGCCAAACAGATGCGATGGCAATGGCGCGGTTGAAGCGCT 1800
QY 1801 GAAAGCACACAACAGTCCAGCATTCGGGTGTTTTGGGCTCGATGCGTGCAGAGCGCT 1860
DB 1801 GAAAGCACACAACAGTCCAGCATTCGGGTGTTTTGGGCTCGATGCGTGCAGAGCGCT 1860
QY 1861 GGCGTGTGAAATCCGGTGCACCTGGGCGGACACCGTACTGAAACGATGCTAACACCGAGC 1920
DB 1861 GGCGTGTGAAATCCGGTGCACCTGGGCGGACACCGTACTGAAACGATGCTAACACCGAGC 1920
QY 1921 GAAAGCGACCTTTGATCTGGGCAAAACCTGGCGATGGTAAAGGTGCGGCTGATGGCAC 1980
DB 1921 GAAAGCGACCTTTGATCTGGGCAAAACCTGGCGATGGTAAAGGTGCGGCTGATGGCAC 1980
QY 1981 CAACTGGAATAATCGACAAACAAAGTGGTCCGGTACCTTATGTTGGCGTAGATAAAGACAA 2040
DB 1981 CAACTGGAATAATCGACAAACAAAGTGGTCCGGTACCTTATGTTGGCGTAGATAAAGACAA 2040
QY 2041 CCTGCTGAATTCAGCAAGAAAGGTACAGTAAAGGAGAGAACTTTTTCATGGAAGTGT 2100
DB 2041 CCTGCTGAATTCAGCAAGAAAGGTACAGTAAAGGAGAGAACTTTTTCATGGAAGTGT 2100
QY 2101 CCCAAATCTGTTGAATTTAGATGGTGTAAATGGGCACAAATTTTCTGTCAGTGGAGA 2160
DB 2101 CCCAAATCTGTTGAATTTAGATGGTGTAAATGGGCACAAATTTTCTGTCAGTGGAGA 2160
QY 2161 GGGTGAAGGTGATGCAACATACGGAATACTTACCTTAAATTTATTTGCACTACTGGAAA 2220
DB 2161 GGGTGAAGGTGATGCAACATACGGAATACTTACCTTAAATTTATTTGCACTACTGGAAA 2220
QY 2221 ACTACTGTTTCAATGGCCAAACACTTGTCACTCTTCTTATGTTGTTCAATGCTTTTC 2280
DB 2221 ACTACTGTTTCAATGGCCAAACACTTGTCACTCTTCTTATGTTGTTCAATGCTTTTC 2280
QY 2281 CGTTTATCCGGATCATATGAAACGGCATGACTTTTTTCAAGAGTGCATGCCCCGAAGTTA 2340

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 2, 2005, 19:32:11 ; Search time 201.5 Seconds
(without alignments)
6426.229 Million cell updates/sec

Title: US-10-649-433C-1
Perfect score: 2180
Sequence: 1 gtttagacgcttatcatcg.....agtttagcggaattgatctg 6729

Scoring table:
OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Word size: 1

Total number of hits satisfying chosen parameters: 565918

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q/cgn2_1/USPTO_spool_p/US10649433/runat 02082005 165026 1831/app_query.fasta_1.6919
-DB=PIR -QFMT=fastan -SUFFIX=oligo.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10649433 @CGN 1 1 332 @runat 02082005 165026 1831 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=100 -DSEBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DSELEXT=7

Database : PIR 79:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	309	14.2	332	1 JGBCG	D-galactose-bindin
2	309	14.2	332	2 B91009	galactose-binding
3	309	14.2	332	2 D85853	galactose-binding
4	284	13.0	360	1 RPCL	lactose operon rep
5	284	13.0	360	4 S58116	lac repressor - sy
6	204	9.4	286	2 T51301	beta-lactamase (EC
7	204	9.4	286	4 S41975	beta-lactamase (EC
8	204	9.4	286	1 PNECP	beta-lactamase (EC
9	185	8.5	286	4 I40905	beta-lactamase (EC
10	185	8.5	286	4 B85529	lac operon transcr
11	183	8.4	360	2 B85529	lac operon transcr
12	183	8.4	360	2 F90678	bla protein - Pseu
13	104	4.8	105	2 JC2566	extended spectrum
14	104	4.8	286	2 S60312	

C	15	90	4.1	149	2	JQ1541	hypothetical 16.9K
	16	86	3.9	286	2	JQ1546	Bla protein - Salm
	17	86	3.9	332	1	S29390	galactose-binding
	18	86	3.9	332	2	AF0781	D-galactose-bindin
C	19	83	3.8	84	1	Q08C7	hypothetical prote
	20	79	3.6	286	2	S30113	beta-lactamase (EC
	21	78	3.6	238	1	JQ1514	green-fluorescent
C	22	70	3.2	91	2	A27354	hypothetical prote
	23	64	2.9	286	2	S60310	extended spectrum
	24	64	2.9	286	2	S60311	beta-lactamase (EC
C	25	62	2.8	88	1	Q08CC8	hypothetical prote
	26	60	2.8	332	1	S15554	D-galactose-bindin
	27	53	2.4	335	2	AG0183	galactose-binding
	28	36	1.7	42	1	Q08C45	hypothetical prote
C	29	32	1.5	293	2	E64874	probable metal-dep
C	30	32	1.5	293	2	F90858	probable enzymes
C	31	32	1.5	293	2	A85761	probable enzymes y
	32	28	1.3	313	2	A48903	beta-lactamase - p
C	33	22	1.0	293	2	AG0653	conserved hypotet
	34	18	0.8	1630	2	A53577	ascites stialoglyco
C	35	17	0.8	54	2	S12504	mb-1 protein - mou
	36	17	0.8	324	2	C82214	galactoside ABC tr
	37	16	0.7	349	1	G64096	D-galactose-bindin
	38	15	0.7	306	1	B45822	beta-lactamase (EC
C	39	15	0.7	507	2	I40767	catalase (EC 1.11.
C	40	14	0.6	231	2	JC7702	NZ-3 antigen - hum
C	41	14	0.6	493	2	A55092	catalase (EC 1.11.
	42	13	0.6	298	2	A41381	beta-lactamase (EC
C	43	13	0.6	539	2	S00067	xylan 1,4-beta-xyl
C	44	12	0.5	63	1	RGECE	regulatory protein
	45	12	0.6	279	2	A24469	beta-lactamase (EC

ALIGNMENTS

RESULT 1

JGBCG

D-galactose-binding periplasmic protein precursor - Escherichia coli (strain K-12)
N:Alternate names: mglB protein
C:Species: Escherichia coli

C>Date: 02-Apr-1982 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A37277; A92319; A94599; A30384; A32653; E64983; A03427; Q00294
R:Hogg, R.W.; Voelker, C.; Von Carlowitz, I.
Mol. Gen. Genet. 229, 453-459, 1991

A>Title: Nucleotide sequence and analysis of the mgl operon of Escherichia coli K12.
A:Reference number: A37277; MUID:92049246; PMID:1719366
A:Accession: A37277

A:Molecule type: DNA

A:Residues: 1-332 <HOG>

A:Cross-references: UNIPROT:P02927; GB:M59444; NID:g146852; PIDN:AAA24169.1; PID:g14685.

A:Experimental source: strain K12

R:Mahoney, W.C.; Hogg, R.W.; Hermodson, M.A.

J. Biol. Chem. 256, 4350-4356, 1981

A>Title: The amino acid sequence of the D-galactose-binding protein from Escherichia coli

A:Reference number: A32319; MUID:81168234; PMID:7012152

A:Accession: A92319

A:Molecule type: protein

A:Residues: 24-178, 'KE', 181-332 <MA1>

A:Experimental source: strain B/r

R:Mahoney, W.C.; Hogg, R.W.; Hermodson, M.A.

submitted to the Atlas, November 1982

A:Reference number: A94599

A:Contents: revisions

A:Accession: A94599

A:Molecule type: protein

A:Residues: 24-332 <MA2>

R:Scholle, A.; Vreemann, J.; Blank, V.; Nold, A.; Boos, W.; Manson, M.D.

Mol. Gen. Genet. 208, 247-253, 1987

A>Title: Sequence of the mglB gene from Escherichia coli K12: comparison of wild-type a

A:Reference number: A30384; MUID:87286407; PMID:3302609

A:Accession: A30384

A:Molecule type: DNA

A:Residues: 1-15 <SCH>
R:Scripture, J.B.; Hogg, R.W.
J. Biol. Chem. 258, 10853-10855, 1983
A:Title: The nucleotide sequences defining the signal peptides of the galactose-binding
A:Reference number: A32653; MUID:83291030; PMID:6885805
A:Accession: A32653
A:Molecule type: DNA
A:Residues: 1-15 <SCR>
R:Vyas, N.K.; Vyas, M.N.; Quioco, F.A.
Proc. Natl. Acad. Sci. U.S.A. 80, 1792-1796, 1983
A:Title: The 3 angstrom resolution structure of a D-galactose-binding protein for transp
A:Reference number: A3942; MUID:83169767; PMID:6340108
A:Contents: annotation; X-ray crystallography, 3.0 angstroms
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: E64983
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-332 <BLAT>
A:CROSS-references: GB:A000304; GB:U00096; NID:gl788470; PIDN:AAC75211.1; PID:gl788473;
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: mglB; mglM
A:Superfamily: D-galactose-binding protein
C:Keywords: calcium binding; chemotaxis; periplasmic space; sugar transport
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-332/Product: D-galactose-binding protein #status experimental <MAT>
F:37,114,175,177,181,234,259,279/Binding site: galactose (Asp, Asn, His, Asp, Arg, Asn, e
F:157,159,161,163,165,228/Binding site: calcium (Asp, Asn, Asp, Gln, Glu) #status e

Alignment Scores:
Pred. No.: 3 98e-308 Length: 332
Score: 309.00 Matches: 309
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.17% Indels: 0
DB: 1 Gaps: 0

US-10-649-433C-1 (1-6729) x JGECG (1-332)

QY 1136 GCTGATCTCGCATTCGATGTAACATCTATAGTACGACGATACTTTATGCTCTAGTG 1195
Db 24 AlaAspThrArgIleGlyValThrIleTyrLysTyrAspAspAsnPheSerValVal 43
QY 1196 CGCAGGCTATTGACGACGATCGGAAGCGCGCCAGATGTTTCAGCTGCTGATGATGAT 1255
Db 44 ArgLysAlaIleGluGlnAspAlaLysAlaAlaProAspValGlnLeuLeuMetAsnAsp 63
QY 1256 TCTCAGATGACAGTCCAAAGCAGAACGATCAGATCGAGTATTGCTGGCCAAAGGGGTG 1315
Db 64 SerGlnAsnAspGlnSerLysGlnAsnAspGlnIleAspValLeuLeuAlaLysGlyVal 83
QY 1316 AAGCAGCTGGCCATCAACTGTTGACCCGCGCAGCTCGGCTACGTTGATTGAGAAAGCG 1375
Db 84 LysAlaLeuAlaIleAsnLeuValAspProAlaAlaAlaGlyThrValIleGluLysAla 103
QY 1376 CGTGGGCAAAAGCTCGCGGTGGTTTCTTCACAAAGACCGTCTCGTAAGCGCTGGAT 1435
Db 104 ArgGlyGlnAsnValProValPhePheAsnLysGluProSerArgLysAlaLeuAsp 123
QY 1436 AGCTACGACAAAGCCCTACTACCTGGCAGTACTCAAAAGAGTCGCGGATATTTCAGGCG 1495
Db 124 SerTyrAspLysAlaTyrTyrValGlyThrAspSerLysGluSerGlyIleIleGlnGly 143
QY 1496 GATTGATGCTAAACATCTGGCGCGGAATCAGGTTGGGATCTGAACAAAGACGGTCAAG 1555
Db 144 AspLeuIleAlaLysHisTrpAlaAlaAsnGlnGlyTyrAspLeuAsnLysAspGlyGln 163
QY 1556 ATTGAGTTCGTACTGCTGAAAAGGTGAAACCGGGCCATCCGGATGTCAGAACGACGTACC 1615

Db 164 IleGlnPheValLeuLeuLysGlyGluProGlyHisProAspAlaGluAlaArgThrThr 183
QY 1616 TACGTGATTAAAGAAATTGAACAGTAAGGATCAAAACTGAACAGTACAGTTACAGTAGATACC 1675
Db 184 TyrValIleLysGluLeuAsnAspLysGlyIleLysThrGluGlnLeuGlnLeuAspThr 203
QY 1676 GCAATGTGGGACACCGCTCAGCGGAAAGATGAAGATGAGCGCTGCTGTCTGGCCCGAAC 1735
Db 204 AlaMetTrpAspThrAlaGlnAlaLysAspLysMetAspAlaTrpLeuSerGlyProAsn 223
QY 1736 GCCAACAAAATCGAAGTGGTTATCCCAACACCATGCGATGCGCAATGCGCGCGTTCGAA 1795
Db 224 AlaAsnLysIleGluValValIleAlaAsnAsnAspAlaMetAlaMetGlyAlaValGlu 243
QY 1796 CGCTGAAAGCACACCAACAGTCACGATTCGGGTTCGGTGGTGGTGGTGGTGGTGGTGG 1855
Db 244 AlaLeuLysAlaHisAsnLysSerSerIleProValPheGlyValAspAlaLeuProGlu 263
QY 1856 CGCTGGCGCTGGTGAATCCGGTGCACTGGCGGACCGTACCTGCTGCTGCTGCTGCTGCTG 1915
Db 264 AlaLeuAlaLeuValLysSerGlyAlaLeuAlaGlyThrValLeuAsnAspAlaAsnAsn 283
QY 1916 CAGCGAAAGCGACCTTTGATCTGCGGAAAACCTGGCGGATGGTAAAGGTGCGGCTGAT 1975
Db 284 GlnAlaLysAlaThrPheAspLeuAlaLysAsnLeuAlaAspGlyLysGlyAlaAlaAsp 303
QY 1976 GGCACCACTGGAAAAATCGACAAAGTGGTGGCGGTACCTTATGTTGGCGTAGATAAA 2035
Db 304 GlyThrAsnTrpLysIleAspAsnLysValValArgValProTyrValGlyValAspLys 323
QY 2036 GACACCTGGCTGAATTCAGCAAGAAA 2062
Db 324 AspAsnLeuAlaGluPheSerLysLys 332

RESULT 2
B31009
galactose-binding transport protein EC3042 [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: B91009
R:Havashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:CROSS-references: UNIPROT:O8X641; GB:BA000007; PIDN:BA36465.1; PID:gl3362511; GSPDB:1-C
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: B91009
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-332 <HAY>
A:CROSS-references: UNIPROT:O8X641; GB:BA000007; PIDN:BA36465.1; PID:gl3362511; GSPDB:1-C
A:Experimental source: strain O157:H7, substrain RIMD 050952
C:Genetics:
A:Gene: EC3042
C:Superfamily: D-galactose-binding protein

Alignment Scores:
Pred. No.: 3 98e-308 Length: 332
Score: 309.00 Matches: 309
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.17% Indels: 0
DB: 2 Gaps: 0

US-10-649-433C-1 (1-6729) x B91009 (1-332)

QY 1136 GCTGATCTCGCATTCGATGTAACATCTATAGTACGACGATACTTTATGCTCTAGTG 1195
Db 24 AlaAspThrArgIleGlyValThrIleTyrLysTyrAspAspAsnPheSerValVal 43
QY 1196 CGCAGGCTATTGACGACGATCGGAAGCGCGCCAGATGTTTCAGCTGCTGATGATGAT 1255
Db 44 ArgLysAlaIleGluGlnAspAlaLysAlaAlaProAspValGlnLeuLeuMetAsnAsp 63
QY 1256 TCTCAGATGACAGTCCAAAGCAGAACGATCAGATCGAGTATTGCTGGCCAAAGGGGTG 1315
Db 64 SerGlnAsnAspGlnSerLysGlnAsnAspGlnIleAspValLeuLeuAlaLysGlyVal 83
QY 1316 AAGCAGCTGGCCATCAACTGTTGACCCGCGCAGCTCGGCTACGTTGATTGAGAAAGCG 1375
Db 84 LysAlaLeuAlaIleAsnLeuValAspProAlaAlaAlaGlyThrValIleGluLysAla 103
QY 1376 CGTGGGCAAAAGCTCGCGGTGGTTTCTTCACAAAGACCGTCTCGTAAGCGCTGGAT 1435
Db 104 ArgGlyGlnAsnValProValPhePheAsnLysGluProSerArgLysAlaLeuAsp 123
QY 1436 AGCTACGACAAAGCCCTACTACCTGGCAGTACTCAAAAGAGTCGCGGATATTTCAGGCG 1495
Db 124 SerTyrAspLysAlaTyrTyrValGlyThrAspSerLysGluSerGlyIleIleGlnGly 143
QY 1496 GATTGATGCTAAACATCTGGCGCGGAATCAGGTTGGGATCTGAACAAAGACGGTCAAG 1555
Db 144 AspLeuIleAlaLysHisTrpAlaAlaAsnGlnGlyTyrAspLeuAsnLysAspGlyGln 163
QY 1556 ATTGAGTTCGTACTGCTGAAAAGGTGAAACCGGGCCATCCGGATGTCAGAACGACGTACC 1615

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2005, 01:28:36 ; Search time 19524 Seconds
(without alignments)
13118.955 Million cell updates/sec

Title: US-10-649-433C-1
Perfect score: 6729
Sequence: 1 gtttagacgcttatcatcg.....agtttagcggaattgatctg 6729

Scoring table: OLIGO_NUC
Gapop_60.0 , Gapext 60.0

Searched: 34239544 seqs, 19032134700 residues

Word size : 0

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*

- 1: gb_est1:*
- 2: gb_est2:*
- 3: gb_hic:*
- 4: gb_est3:*
- 5: gb_est4:*
- 6: gb_est5:*
- 7: gb_est6:*
- 8: gb_gsa1:*
- 9: gb_gsa2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	864	12.8	1070	1 AJ281552	4A3A-P6F1
C 2	788	11.7	841	1 AL042026	DKFZP434E
C 3	770	11.4	865	7 CK125894	BES182411
C 4	742	11.0	793	5 BQ751655	EST632218
C 5	742	11.0	806	5 BQ751220	EST631783
C 6	742	11.0	827	7 CN823902	Oa splbn
C 7	741	11.0	758	9 CL422788	AE0544 Sa
C 8	736	10.9	856	7 CN823189	Oa splbn
C 9	732	10.9	928	7 CO487414	CO0227.B7
C 10	723	10.7	1004	1 AJ281480	4A3A-P4G8
C 11	722	10.7	789	6 CD280920	G44224.42
C 12	717	10.7	769	7 CK118014	218n06.p1
C 13	706	10.5	759	6 CD279661	G43818.35
C 14	699	10.4	780	5 BQ825693	1030129B0
C 15	699	10.4	846	7 CV468077	est_l van
C 16	690	10.3	741	6 CD279174	G44221.83
C 17	679	10.1	730	6 CD281097	G44224.38
C 18	678	10.1	729	6 CD279546	G44222.81
C 19	678	10.1	800	1 AJ281449	4A3A-P4D5
C 20	677	10.1	728	6 CD279322	G44223.13
C 21	675	10.0	726	6 CD281811	G44271.30
C 22	674	10.0	954	1 AL044364	DKFZP434C
C 23	673	10.0	710	7 CK394569	hgada5A0
C 24	671	10.0	733	8 BZ049372	jnr50b10.

C	25	671	10.0	779	7	CK394523	hgada4D0
	26	666	9.9	727	7	CR791051	DKFZP469B
	27	665	9.9	794	5	BQ751936	EST632499
	28	657	9.8	708	6	CD281616	CD281616
	29	655	9.7	706	7	CR543261	DKFZP470P
	30	652	9.7	703	6	CD279544	G44222.73
	31	652	9.7	832	7	CN822433	Oa splbn
	32	649	9.6	700	5	BQ660293	H101N05w
	33	648	9.6	699	5	BQ659852	HF01022w
	34	648	9.6	700	6	CA305674	EST000022
	35	648	9.6	819	6	CD849375	CV6nd0008
	36	644	9.6	697	5	BQ660197	H101G17w
	37	644	9.6	697	5	BQ660261	H101L03w
	38	644	9.6	793	7	CK122681	BES182410
	39	641	9.5	641	4	BJ063872	BJ063872
	40	640	9.5	1013	4	BM438846	IP4vR0015
	41	638	9.5	697	7	CK739326	EA0282 Da
	42	638	9.5	832	4	BG923768	602825893
	43	636	9.5	769	7	CV224987	CS hyp 24
	44	636	9.5	786	7	CN823164	Oa splbn
	45	636	9.5	806	4	BJ684280	BJ684280

ALIGNMENTS

RESULT 1
AJ281552/c
LOCUS
DEFINITION
4A3A-P6F11-F Anopheles gambiae immune competent 4A3A Anopheles gambiae cDNA clone 4A3A-P6F11, mRNA sequence.
ACCESSION
AJ281552
VERSION
AJ281552.1 GI:6929432
KEYWORDS
EST.
SOURCE
Anopheles gambiae (African malaria mosquito)
ORGANISM
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
REFERENCE
1 (bases 1 to 1070)
Dimopoulos,G., Casavantz,T.L., Chang,S., Scheetz,T., Roberts,C., Donohue,M., Schultz,J., Benes,V., Bork,P., Ansorge,W., Soares,M.B. and Kafatos,F.C.
Anopheles gambiae pilot gene discovery project: identification of mosquito innate immunity genes from expressed sequence tags generated from immune-competent cell lines
Proc. Natl. Acad. Sci. U.S.A. 97 (12), 6619-6624 (2000)
10841561
20300950
COMMENT
Contact: Dimopoulos G
Fotis C. Kafatos laboratory
European Molecular Biology Laboratory
Meyerhofstrasse 1, 69117 Heidelberg, Germany.
Location/Qualifiers
1. .1070
/organism="Anopheles gambiae"
/mol_type="mRNA"
/strain="4A r/r"
/db_xref="taxon:7165"
/clone="4A3A-P6F11"
/cell_line="immune competent 4A3A"
/lab_host="E. coli DH10B"
/note="Vector: pRTT3D-Pac (Pharmacia) with a modified polylinker; Site 1: EcoRI; Site 2: NotI; sequenced from forward priming site which reads from the 3' end of the cDNA. The 4A3A is a directionally cloned and normalized cDNA library that was constructed from the 4A3A cell line oligo-T primed cDNA according to: Bonaldo, Lennon & Soares (1996) : Normalization and Subtraction: two approaches to Facilitate Gene Discovery, Genome Research 6, 791-806."

FEATURES
source

ORIGIN

Query Match		12.8%;	Score 864;	DB 1;	Length 1070;
Best Local Similarity		99.8%;	Pred. No. 0;		
Matches 964;		Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
QY	4064	CTTTCGGCTGGCTGGTTTATTGCTGTATATAATCTCGAGCCGGTAGCGTGGTCTCGCGG	4123		
DB	966	CTTTCGGCTGGCTGGTTTATTGCTGTATATAATCTCGAGCCGGTAGCGTGGTCTCGCGG	907		
QY	4124	TATCATTTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACAGAC	4183		
DB	906	TATCATTTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACAGAC	847		
QY	4184	GGGGAGTCAGCAACTATCGATGAACGGAATAGACAGATCGCTGAGATAGTGCCTCACT	4243		
DB	846	GGGGAGTCAGCAACTATCGATGGACGAATAGACAGATCGCTGAGATAGTGCCTCACT	787		
QY	4244	GATTAAAGCACTGGTAACCTGTGACACCAAGTTTACTCATATATACCTTTAGATTGATTAAA	4303		
DB	786	GATTAAAGCACTGGTAACCTGTGACACCAAGTTTACTCATATATACCTTTAGATTGATTAAA	727		
QY	4304	ACTTCATTTTAAATTTAAAGGATCTAGGTGAAGATCCTTTTGTGATAATCTCATGACCA	4363		
DB	726	ACTTCATTTTAAATTTAAAGGATCTAGGTGAAGATCCTTTTGTGATAATCTCATGACCA	667		
QY	4364	AATCCCTTAACTGAGTTTCTGTTCCACTGAGCGTCAGACCCGTAGAAAAGATCAAAGG	4423		
DB	666	AATCCCTTAACTGAGTTTCTGTTCCACTGAGCGTCAGACCCGTAGAAAAGATCAAAGG	607		
QY	4424	ATCTTCTTGAGATCCTTTTCTCGCGCTAATCTGCTGCTTGGCAACAAAAACCAACC	4483		
DB	606	ATCTTCTTGAGATCCTTTTCTCGCGCTAATCTGCTGCTTGGCAACAAAAACCAACC	547		
QY	4484	GCTACAGCGGTGGTTGTTGCGCGATCAAGAGCTACCAACTCTTTTTCGAAGGTTAAC	4543		
DB	546	GCTACAGCGGTGGTTGTTGCGCGATCAAGAGCTACCAACTCTTTTTCGAAGGTTAAC	487		
QY	4544	TGGCTTCAGCAGAGCGAGATACCAAAATACTGTCTTCTAGTGTAGCCGTAGTTAGGCCA	4603		
DB	486	TGGCTTCAGCAGAGCGAGATACCAAAATACTGTCTTCTAGTGTAGCCGTAGTTAGGCCA	427		
QY	4604	CCACTTCAAGAACTCTGTAGCAGCCGCTACATACCTCGCTCTGCTAACTCTTTACAGT	4663		
DB	426	CCACTTCAAGAACTCTGTAGCAGCCGCTACATACCTCGCTCTGCTAACTCTTTACAGT	367		
QY	4664	GGCTGCTGCCAGTGGCGATAAGTCTGTCTTACCGGTTGGACTCAAGACGATAGTTACC	4723		
DB	366	GGCTGCTGCCAGTGGCGATAAGTCTGTCTTACCGGTTGGACTCAAGACGATAGTTACC	307		
QY	4724	GGATAAGGCGAGCGGTTCGGCTGAAACGGGGGGTTGCTGCACACAGCCCGCTTGGAGCG	4783		
DB	306	GGATAAGGCGAGCGGTTCGGCTGAAACGGGGGGTTGCTGCACACAGCCCGCTTGGAGCG	247		
QY	4784	AACGACCTACACCGAACTGAGATACCTACAGCGTGAAGCTATGAGAAAGCGCCAGCTTCC	4843		
DB	246	AACGACCTACACCGAACTGAGATACCTACAGCGTGAAGCTATGAGAAAGCGCCAGCTTCC	187		
QY	4844	CGAAGGGAAGAGCGGACAGGTATCCGGTAAGCGGCGAGGGTCGGAACAGGAGACGCGAC	4903		
DB	186	CGAAGGGAAGAGCGGACAGGTATCCGGTAAGCGGCGAGGGTCGGAACAGGAGACGCGAC	127		
QY	4904	GAGGAGCTTCCAGGGGGGAAACGCTTGTTATCTTTATAGTCTGTCGGGTTTCGCCACT	4963		
DB	126	GAGGAGCTTCCAGGGGGGAAACGCTTGTTATCTTTATAGTCTGTCGGGTTTCGCCACT	67		
QY	4964	CTGACTTGAGCGTCGATTTTGTGATGCTCGTCAGGGGGCGGAGCCTATGAAAAACGC	5023		
DB	66	CTGACTTGAGCGTCGATTTTGTGATGCTCGTCAGGGGGCGGAGCCTATGAAAAACGC	7		
QY	5024	CAGCAA 5029			
DB	6	CAGCAA 1			

RESULT 2	AL042026	AL042026	DEFINITION	DKFZp434E111_r1 434 (synonym: htes3) Homo sapiens cDNA clone	841 bp	mRNA	linear	EST 04-SEP-2003
LOCUS	DKFZp434E111	5'	EST	DKFZp434E111	5'	mRNA	sequence.	
ACCESSION	AL042026	AL042026	VERSION	AL042026.1	GI:5421372	EST		
KEYWORDS	EST		SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens		REFERENCE	1 (bases 1 to 841)				
AUTHORS	Poustka, A., Klein, M., Mewes, H.W., Gassenhuber, J. and Wiemann, S.		TITLE	EST (Poustka, et al.)				
JOURNAL	Unpublished (1999)		COMMENT	Contact: MIPS				
FEATURES	source							
	1..841							
	/organism="Homo sapiens"							
	/mol_type="mRNA"							
	/db_xref="taxon:9606"							
	/clone="DKFZp434E111"							
	/tissue_type="testis"							
	/dev_stage="adult"							
	/lab_host="DH10B"							
	/clone_lib="434"							
	/note="Vector: pSport1; Site_1: NotI; Site_2: SalI"							
ORIGIN								
Query Match	11.7%;	Score 788;	DB 1;	Length 841;				
Best Local Similarity	99.9%;	Pred. No. 0;						
Matches 838;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;				
QY	3307	TTTGTGTTATTTTCTAAATACATTCACAAATATGTATCCGCTCATGAGACAATAACCTGAT	3366					
DB	3	TTTGTGTTATTTTCTAAATACATTCACAAATATGTATCCGCTCATGAGACAATAACCTGAT	62					
QY	3367	AAATGCTTCAATAATATTGAAAAAGAGAGATGAGTATGAGTATTCACATTTCCGTGTCGCC	3426					
DB	63	AAATGCTTCAATAATATTGAAAAAGAGAGATGAGTATGAGTATTCACATTTCCGTGTCGCC	122					
QY	3427	TTATTCCTTTTTCGGCATTTCCTCTCTGTTTTCCTCACCACCAAGACGTGGTGA	3486					
DB	123	TTATTCCTTTTTCGGCATTTCCTCTCTGTTTTCCTCACCACCAAGACGTGGTGA	182					
QY	3487	AAGTAAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTCATCATCAACTGGATCTCA	3546					
DB	193	AAGTAAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTCATCATCAACTGGATCTCA	242					
QY	3547	ACAGCGGTAAAGATCCTTGAGAGTTTTCGCCCGGAAGACGTTTTCATATGATGACACTT	3606					
DB	243	ACAGCGGTAAAGATCCTTGAGAGTTTTCGCCCGGAAGACGTTTTCATATGATGACACTT	302					
QY	3607	TTAAGTTCTGCTATGTCGGCGGTATTATCCCGTGTGACGCCGGCAAGCAACTCG	3666					
DB	303	TTAAGTTCTGCTATGTCGGCGGTATTATCCCGTGTGACGCCGGCAAGCAACTCG	362					
QY	3667	GTGCGCGCATACACTATTCTCAGATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGC	3726					
DB	363	GTGCGCGCATACACTATTCTCAGATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGC	422					

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 2, 2005, 19:25:21 ; Search time 1028 Seconds
(without alignments)
6703.854 Million cell updates/sec

Title: US-10-649-433C-1

Perfect score: 2180

Sequence: 1 gtttagacgttatcatcga.....agtttagcggaattgatctg 6729

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1612378 seqs, 512079187 residues

Word size: 1

Total number of hits satisfying chosen parameters: 3224408

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters: -DEV-xlp

-Q/cgn2_1/USFTO_spool_P/US10649433/runat_02082005_165025_1820/app_query.fasta_1.6919
-DB=UniProt -QFMT=fastan -SUFFIX=oligo.rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10649433 @CGN 1 1 1271 @runat_02082005_165025_1820 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : UniProt 03.*

1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	309	14.2	332	1 DGAL_ECOLI	P02927 escherichia
2	309	14.2	332	2 Q8X641	Q8X641 escherichia
3	284	13.0	360	1 LACI_ECOLI	P03023 escherichia
4	222	10.2	286	2 Q91UY8	Q91UY8 plasmid pww
5	204	9.4	286	2 Q38058	Q38058 bacterioph
6	204	9.4	286	2 Q00626	Q00626 staphyloc
7	204	9.4	286	2 Q79CL6	Q79CL6 methylobac
8	204	9.4	286	2 Q79DR3	Q79DR3 escherichia
9	203	9.3	285	2 Q6XB67	Q6XB67 hordeum vul
10	185	8.5	286	1 BLAT_ECOLI	P62594 salmonella
11	185	8.5	286	1 BLAT_SALTI	P62594 salmonella
12	185	8.5	286	2 Q6A253	Q6A253 haemophilus
13	185	8.5	286	2 Q6LBN9	Q6LBN9 pseudomonas
14	185	8.5	286	2 Q6LCV6	Q6LCV6 neisseria g
15	185	8.5	286	2 Q6TMH1	Q6TMH1 streptococ
16	185	8.5	286	2 Q6W9J1	Q6W9J1 enterobacte

17	185	8.5	286	2 Q6WRX2	Q6WRX2 zymomonas m
18	185	8.5	286	2 Q6WZD4	Q6WZD4 acinetobact
19	185	8.5	286	2 Q7B3X5	Q7B3X5 citrobacter
20	185	8.5	286	2 Q7B899	Q7B899 klebsiella
21	185	8.5	286	2 Q7BP57	Q7BP57 shigella fl
22	185	8.5	286	2 Q7BR75	Q7BR75 neisseria m
23	185	8.5	286	2 Q7DFY3	Q7DFY3 salmonella
24	185	8.5	286	2 Q7DHD3	Q7DHD3 serratia ma
25	185	8.5	286	2 Q799Y1	Q799Y1 plasmid ppv
26	185	8.5	286	2 Q7BVP8	Q7BVP8 plasmid pet
27	183	8.4	360	2 Q7AH57	Q7AH57 escherichia
28	183	8.4	360	2 Q8X684	Q8X684 escherichia
29	181	8.3	225	2 Q38212	Q38212 bacterioph
30	181	8.3	281	2 Q6Q1V0	Q6Q1V0 serratia ma
31	181	8.3	282	2 Q6Q1V1	Q6Q1V1 serratia ma
32	181	8.3	286	2 Q8KSD3	Q8KSD3 klebsiella
33	181	8.3	286	2 Q93A77	Q93A77 escherichia
34	181	8.3	286	2 Q93G13	Q93G13 klebsiella
35	181	8.3	286	2 Q6SQJ9	Q6SQJ9 streptococ
36	181	8.3	286	2 Q6UVM7	Q6UVM7 acinetobact
37	181	8.3	286	2 Q6W7J4	Q6W7J4 escherichia
38	181	8.3	286	2 Q8GA85	Q8GA85 escherichia
39	181	8.3	286	2 Q8EYX1	Q8EYX1 escherichia
40	181	8.3	286	2 Q9K582	Q9K582 klebsiella
41	181	8.3	286	2 Q9R771	Q9R771 klebsiella
42	181	8.3	286	2 Q9RMS4	Q9RMS4 escherichia
43	181	8.3	286	2 Q9RN48	Q9RN48 escherichia
44	181	8.3	286	2 Q9S679	Q9S679 escherichia
45	180	8.3	286	2 Q933Z8	Q933Z8 escherichia

ALIGNMENTS

RESULT 1

DGAL_ECOLI
ID DGAL_ECOLI STANDARD; PRT; 332 AA.
AC P02927; P17775;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 25-JAN-2005 (Rel. 46, Last annotation update)
DE D-galactose-binding periplasmic protein precursor (GBP) (D-galactose/
D-glucose binding protein) (GGBP).
GN Name=mgIB; OrderedLocustNames=b2150, c2684, SF2235, S2364;
OS Escherichia coli,
OS Escherichia coli O6, and
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 217992, 623;
RN [1]_TaxID=562, 217992, 623;
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=K12;
RX MEDLINE=92049246; PubMed=1719366;
RA Hogg R.W., Voelker C., von Carlowitz I.;
RT "Nucleotide sequence and analysis of the mgI operon of Escherichia coli K12".
RL Mol. Gen. Genet. 229:453-459(1991).
RN [2]
RP SEQUENCE FROM N.A., AND MUTAGENESIS OF GLY-74.
RC SPECIES=E.coli; STRAIN=K12;
RX MEDLINE=87286407; PubMed=3302609;
RA Scholle A., Vreemann J., Blank V., Nold A., Boos W., Manson M.D.;
RT "Sequence of the mgIB gene from Escherichia coli K12: comparison of wild-type and mutant galactose chemoreceptors".
RL Mol. Gen. Genet. 208:247-253(1987).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=K12 / BHB2600;
RA Richterich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K., Church G.M.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.

RC SPECIES=E.coli; STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.,
RT "The complete genome sequence of *Escherichia coli* K-12.";
RL Science 277:1453-1474 (1997).
RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=O6:H1 / CFT073 / ATCC 700928 / UPEC;
RX MEDLINE=22388234; PubMed=12471157; DOI=10.1073/pnas.252529799;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic *Escherichia coli*.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
RN [6]
RP SEQUENCE OF 24-332.
RC SPECIES=E.coli;
RX MEDLINE=81168234; PubMed=7012152;
RA Mahoney W.C., Hogg R.W., Hermodson M.A.;
RT "The amino acid sequence of the D-galactose-binding protein from
Escherichia coli B/r.";
RL J. Biol. Chem. 256:4350-4356 (1981).
RN [7]
RP REVISIONS TO 179-180.
RC SPECIES=E.coli;
RA Mahoney W.C., Hogg R.W., Hermodson M.A.;
RL Submitted (NOV-1982) to the PIR data bank.
RN [8]
RP SEQUENCE OF 1-58 FROM N.A.
RC SPECIES=E.coli;
RX MEDLINE=83291030; PubMed=6885805;
RA Scripture J.B., Hogg R.W.;
RT "The nucleotide sequences defining the signal peptides of the
galactose-binding protein and the arabinose-binding protein.";
RL J. Biol. Chem. 258:10853-10855 (1983).
RN [9]
RP SEQUENCE OF 24-47.
RC SPECIES=E.coli; STRAIN=K12 / EMG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
in the genome of *Escherichia coli* K-12.";
RL Electrophoresis 18:1259-1313 (1997).
RN [10]
RP SEQUENCE FROM N.A.
RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
RX MEDLINE=2272406; PubMed=12384590; DOI=10.1093/nar/gkf566;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding X., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity
through comparison with genomes of *Escherichia coli* K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441 (2002).
RN [11]
RP SEQUENCE FROM N.A.
RC SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA DOI=10.1128/IAI.71.5.2775-2786.2003;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of *Shigella*
flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786 (2003).
RN [12]

RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RC SPECIES=E.coli;
RX MEDLINE=83169767; PubMed=6340108;
RA Vyas N.K., Vyas M.N., Quicho F.A.;
RT "The 3-A resolution structure of a D-galactose-binding protein for
transport and chemotaxis in *Escherichia coli*.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:1792-1796 (1983).
RN [13]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RC SPECIES=E.coli;
RX MEDLINE=87258173; PubMed=3600760; DOI=10.1038/327635a0;
RA Vyas N.K., Vyas M.N., Quicho F.A.;
RT "A novel calcium binding site in the galactose-binding protein of
bacterial transport and chemotaxis.";
RL Nature 327:635-638 (1987).
RN [14]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RC SPECIES=E.coli;
RX MEDLINE=89058638; PubMed=3057628;
RA Vyas N.K., Vyas M.N., Quicho F.A.;
RT "Sugar and signal-transducer binding sites of the *Escherichia coli*
galactose chemoreceptor protein.";
RL Science 242:1290-1295 (1988).
CC -!- FUNCTION: This protein is involved in the active transport of
galactose and glucose. It plays a role in the chemotaxis towards
the two sugars by interacting with the trg chemoreceptor.
CC -!- SUBCELLULAR LOCATION: Periplasmic.
CC -!- DOMAIN: The calcium-binding site is structurally similar to that
of EF-hand proteins, but is in two parts, with the last calcium
ligand provided by Glu-228.
CC -!- SIMILARITY: Belongs to the bacterial extracellular solute-binding
protein family 2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; M59444; AAA241169.1; -.
CC EMBL; U00007; AAA60523.1; -.
CC EMBL; X05646; CAA29132.1; -.
CC EMBL; U00096; AAC75211.1; -.
CC EMBL; AE016763; AAN81140.1; -.
CC EMBL; K00419; AAA24172.1; -.
CC EMBL; AE015239; AAN43756.1; -.
CC EMBL; AE016985; AAP17573.1; -.
CC PIR; A37277; JGECG.
CC PDB; 1GLG; X-ray; @=24-332.
CC PDB; 2GBP; X-ray; @=24-332.
CC SWISS-2DPAGE; P02927; COLI.
CC EC02DRASE; C029.8; 6TH EDITION.
CC EchoBASE; EB0588; -.
CC Ecogene; EGI0593; mgIB.
CC InterPro; IPR001761; PeriPlaBP/LacI.
CC Pfam; PF00532; PeriPla_BP_1; 1.
KW 3D-structure; Calcium-binding; Chemotaxis; Complete proteome;
KW Direct protein sequencing; Periplasmic; Signal; Sugar transport;
KW Transport.
FT SIGNAL 1 23
FT CHAIN 24 332 D-galactose-binding periplasmic protein.
FT CA_BIND 157 165 First part of site.
FT CA_BIND 227 228 Second part of site.
FT SITE 97 97 Interacts with membrane-bound trg signal
transducer.
FT MUTAGEN 97 97 G->D: Improductive interaction with trg.
FT STRAND 26 32
FT TURN 35 36
FT HELIX 38 51
FT TURN 52 53
FT TURN 55 55

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 2, 2005, 18:10:32 ; Search time 802 Seconds
(without alignments)
6490.051 Million cell updates/sec

Title: US-10-649-433C-1

Perfect score: 11920

Sequence: 1 gtttagacgcttatcatgca.....agttagcggaattgatctg 6729

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q/cn2_1/USPTO spool_p/US10649433/runat_02082005_164944_1513/app_query.fasta_1.6919
-DB=A Geneseq -QFM=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10649433 @CN 1 1 977 @runat_02082005_164944_1513 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A Geneseq_16Dec04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8639.5	72.5	1967	AB882793	Amino aci
2	4004	33.6	799	ABr43385	FLIPglu f
3	3273	27.1	2606	ADe52217	pG1EN-EH3
4	3273	27.1	2606	ADe52214	pG1EN-EH3
5	3086.5	25.9	2761	ADh11249	Vertebrat
6	3083.5	25.6	3614	ADh11258	Vertebrat
7	3062.5	25.7	766	ADh11236	Vertebrat
8	3058	25.7	1221	AAR52699	Sequence
9	3024	25.1	2732	ADh11251	Vertebrat
10	3004	24.9	1310	ADc22809	Human G p

C	11	3004	24.9	1310	7	ADh14282	Adh14282 Vector pC
	12	2946.5	24.7	864	7	ADh45227	Modified
	13	2902	24.3	1293	7	ADc22811	Human G p
	14	2902	24.3	1293	7	ADh14284	Vector pC
C	15	2901.5	24.1	2458	2	AAR07640	Deducted p
C	16	2901.5	24.1	2458	2	AAR04031	Full leng
	17	2895.5	24.3	1277	2	AAR52701	Plasmid p
C	18	2890.5	24.0	2363	7	ADG15745	Novel vac
	19	2871	24.1	1277	2	AAR52700	Plasmid p
C	20	2842	23.6	1273	7	ADc22808	Human G p
C	21	2842	23.6	1273	7	ADh14281	Vector pC
	22	2824	23.7	1279	7	ADc22812	Human G p
	23	2824	23.7	1279	7	ADh14285	Vector pC
C	24	2819	23.4	1286	7	ADc22810	Human G p
C	25	2819	23.4	1286	7	ADh14283	Vector pC
C	26	2802	23.2	2050	2	AAR07641	Deducted s
	27	2773.5	23.3	1284	7	ADc22813	Human G p
	28	2773.5	23.3	1284	7	ADh14286	Vector pC
C	29	2709	22.7	1088	2	AAR8636	Plasmid p
	30	2707.5	22.5	2037	2	AAR04032	Full leng
C	31	2675.5	22.4	1088	2	AAR8635	Plasmid p
	32	2626	22.0	2234	2	ADh11230	Vertebrat
	33	2618.5	22.0	1088	2	AAR8634	Plasmid p
	34	2617	22.0	763	6	ABR43387	FLIPrbs f
	35	2592	21.7	2319	2	ADh11228	Vertebrat
	36	2586	21.7	1699	2	ADh11378	Vertebrat
	37	2486.5	20.9	740	4	AAB99360	Human H-R
	38	2486.5	20.9	740	4	AAB99357	Human H-R
	39	2486.5	20.9	740	5	ABB06725	Human H-R
	40	2486.5	20.9	740	5	ABB06728	Human H-R
	41	2486.5	20.9	740	5	ABB06738	Human H-R
	42	2486.5	20.9	740	5	ABB06735	Human H-R
	43	2486.5	20.9	740	5	ABB06730	Human H-R
	44	2486.5	20.9	764	5	ABB06740	Human H-R
	45	2486.5	20.9	764	5	ABB06740	Human H-R

ALIGNMENTS

RESULT 1

AB882793

ID AB882793 standard; protein; 1967 AA.

XX AC

XX AB882793;

XX AC

XX DT 18-MAR-2003 (first entry)

XX DE Amino acid sequence of plasmid pTrcCFRET3.

XX KW Fluorescent; fluorescence resonance energy transfer; FRET; tumour;

XX KW estrogen; tamoxifen; therapy; pTrcCFRET3.

XX OS Synthetic.

XX FH Key

XX FT Misc-difference 1..1967

XX FT /note= "Xaa are residues encoded by internal stop codons"

XX PN WO200290987-A2.

XX PD 14-NOV-2002.

XX PF 10-MAY-2002; 2002WO-GB002183.

XX PR 10-MAY-2001; 2001GB-00011459.

XX PA (ISIS-) ISIS INNOVATION LTD.

XX PI Fricker MD, Vaux DJT;

XX XX WPI; 2003-129191/12.

XX DR N-PSDB; ABZ23939.

XX XX

PT Probe, useful e.g. for medical diagnosis, and detection of pollutants in
 PT water systems and contaminants in foodstuffs, has target binding site
 PT moiety and mimic moiety attached to two fluorescent polypeptides, and
 PT linker.

XX Disclosure; Fig 5A-H; 50pp; English.

CC The invention relates to a probe (I) comprising a target binding site
 CC moiety attached to a first fluorescent polypeptide, a mimic moiety
 CC capable of binding to the target binding site moiety and attached to a
 CC second fluorescent polypeptide, and a linker connecting the two
 CC fluorescent polypeptides and which allows the distance between the
 CC fluorescent polypeptides to vary. (I), or a cell harbouring (I), or a
 CC polynucleotide which encodes (I), or a sensor (V) comprising (I), a light
 CC source which is capable of exciting the probe, and a detecting which is
 CC capable of measuring the amount of fluorescence resonance energy transfer
 CC (FRET) from the probe are useful for detecting the presence or absence of
 CC a target substance in a test sample, and for identifying an inhibitor of
 CC binding between two substances, where the two substances would bind to
 CC each other in the absence of an inhibitor (see AB223939 for a description
 CC of the varied uses of the probe of the invention). The present sequence
 CC represents the amino acid sequence of plasmid pTrcCFRET3

XX Sequence 1967 AA;

Alignment Scores:

Pred. No.: 0 Length: 1967
 Score: 8639.50 Matches: 1701
 Percent Similarity: 81.30% Conservative: 21
 Best Local Similarity: 80.31% Mismatches: 76
 Query Match: 72.48% Indels: 320
 Db: 6 Gaps: 8

US-10-649-433C-1 (1-6729) x ABB82793 (1-1967)

QY	413	ATGTTGAGCAAGGGGAGGAGCTGTTCAACGGGGTGTGCCATCCTCGTGGAGCTGGAC	472
Db	4	MetValSerLysGlyGluGluLeuPheThrGlyValValProIleLeuValGluLeuAsp	23
QY	473	GGCGAGCTAAACGGCCACAAAGTTACGCTGTCCGGCGAGGGCGAGGGGATGCCACCTAC	532
Db	24	GlyAspValAsnGlyHisLysPheSerValSerGlyGlyGluGlyGluAlaThrTyr	43
QY	533	GGCAAGCTGACCTGAAGTTGATCTGACACCGGCAAGCTGCCGTGCCCTGGCCACC	592
Db	44	GlyLysLeuThrLeuLysPheIleCysThrThrGlyLysLeuProValProTrpProThr	63
QY	593	CTCGTGACCACTTCGGCTACGGCTGAGTGTCTGCGCGCTACCCCGCTACCCGACCATGAAG	652
Db	64	LeuValThrLeuThrTrpGlyValGlnCysPheSerArgTyrProAspHisMetLys	83
QY	653	CAGCACGACTTCTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTC	712
Db	84	GlnHisAspPhePheLysSerAlaMetProGluGlyTyrValGlnGluArgThrIlePhe	103
QY	713	TTCAAGGACGAGGCAACTAAGACCGCGCGAGGTGAAGTTCGAGGGCGACCCCTG	772
Db	104	PheLysAspAspGlyAsnTyrLysThrArgAlaGluValLysPheGluGlyAspThrLeu	123
QY	773	GTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGAGCGCAACATCTCGGGCAC	832
Db	124	ValAsnArgIleGluLeuLysGlyIleAspPheLysGluAspGlyAsnIleLeuGlyHis	143
QY	833	AACTGGAGTACAATCAACAGCCCAACAGTCTATATCATGGCCGACAGCAGAGAAC	892
Db	144	LysLeuGluTyrAsnTyrIleSerHisAsnValTyrIleThrAlaAspLysGlnLysAsn	163
QY	893	GGCATCAGGTGAATTCAGATCCGCACACATCGAGGACGGCAGCGTGCAGCTCGCC	952
Db	164	GlyIleLysAlaAsnPheLysIleArgHisAsnIleGluAspGlySerValGlnLeuAla	183
QY	953	GACCACTACGACAGAACACCCCATCGCGACGGCCCGTGTCTGTCTGCCGACCAACAC	1012
Db			

Db	184	AspHisTyrGlnGlnAsnThrProIleGlyAspGlyProValLeuLeuProAspAsnHis	203
QY	1013	TACCTGAGCTACCACTCGCCCTGAGCAAAACCCCAACGAGACGCGATCATCATGTGC	1072
Db	204	TyrLeuSerThrGlnSerAlaLeuSerLysAspProAsnGluLysArgAspHisMetVal	223
QY	1073	CTGCTGGAGTTCTGTCGACCGCGCC-----GGGATCACTCTC	1108
Db	224	LeuLeuGluPheValThrAlaAlaLeuGlnSerSerGlyGlyGlyGlySerThrMet	243
QY	1109	GGCATGGACGAG-----CTGTACAAAGACTAGTGTCTGATACCTCGCATT	1150
Db	244	GlyGlySerHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis	263
QY	1151	GGTGTAACTATATAAGTACGACGATAAATTTATGCTGTAGTCGCGCAAGGCTATTGAG	1210
Db	264	GlyArgAspLeuTyrAspAspAsp-----	272
QY	1211	CAAGATGGAAAGCGCGCCAGATGTTTCAGCTGTGATGATGATTCAGATGATCACCAG	1270
Db	272	-----	272
QY	1271	TCCAAGCAGAACGATCAGATCGATATTGCTGGCCAAAGGGGTGAAGGCACTGGCCATC	1330
Db	272	-----	272
QY	1331	AACCTGGTTGACCGCGCAGCTGCGGGTACGGGTAGTGTGAGAAAGCGCGTGGCAAAACGTG	1390
Db	272	-----	272
QY	1391	CGGTGGTTTTCTTCAACAAAGAACCGTCTCGTAAGGGCGCTGATAGTACGACAAAGCC	1450
Db	272	-----	272
QY	1451	TACTACGTTGGCACTGACTCAAAAGAGTCCGGCATTTATTCAAGGCGATTTGATTCCTAAA	1510
Db	273	-----	273
QY	1511	CACCTGGCGGCGAATTCAGGGTTGGGATCTGAACAAAGCGGTTCAGTTTCGATCTGCTG	1570
Db	274	His-----ArgTrpIleArg	278
QY	1571	CTGAAGGTGAACCGGGCCATCCGGATGCAGAGACGATACCATTCATCGTATTAAAGAA	1630
Db	279	ProArgGlySerSerGly-----	284
QY	1631	TTGAACGATAAAGGCATCAAAACTGAACAGTTACAGTTAGATACCGCAATGTGGGACACC	1690
Db	284	-----	284
QY	1691	GCTCAGGCGAAAGATAAGATGGACGCTGGCTGTCTGGCCCGAAGCGCAACAAATCGAA	1750
Db	284	-----	284
QY	1751	GTGGTTATCGCCAAACAGATCGATGGCAATGGCGCGGTGGAAGCGCTGAAAGCACAC	1810
Db	284	-----	284
QY	1811	AACAAGTCCAGCATTCGCGTGTGGCGTGTGGCGTGTGGCGTGTGGCGTGTGGCGTGTG	1870
Db	284	-----	284
QY	1871	AAATCCGCTGCACTGGCGGGCACCGTACTGAACGATGCTAAACACCGCGGAAAGCGACC	1930
Db	285	-----GlyGlyGlySerGly-----	289
QY	1931	TTTGATCTGGCGGAAAAACCTCGCGCATGTGTAAGTGGCGGTGATGGCACCACTGGAAA	1990
Db	290	-----GlyGlyGlySerGlyGly-----	296
QY	1991	ATCGACAAAGAGTGTCCCGGTACCTTTATGTTGGCGTAGATAAAGACCAACCTGGCTGAA	2050
Db	296	-----	296

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 2, 2005, 18:20:38 ; Search time 160 Seconds
(without alignments)
6278.919 Million cell updates/sec

Title: US-10-649-433C-1

Perfect score: 11920

Sequence: 1 gtttagacgcttatcatcg.....agtttagcggaattgatctg 6729

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 1027090

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/cgn2_1/USPTO spool_p/US10649433/runat_02082005_164946_1547/app_query.fasta_1.6919
-DB=issued_Patents_AA -Qfmt=fastan -SUFFIX=ra1 -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10649433 @CGN 1 1 178 @runat_02082005_164946_1547 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: Issued Patents AA:*
- 2: /cgn2_6/ptodata/1/iaa/5A COMB pep:*
- 3: /cgn2_6/ptodata/1/iaa/5B COMB pep:*
- 4: /cgn2_6/ptodata/1/iaa/6A COMB pep:*
- 5: /cgn2_6/ptodata/1/iaa/6B COMB pep:*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	3004	24.9	1310	4	US-09-170-496D-290
C 2	3004	24.9	1310	4	US-09-364-425B-55
C 3	2902	24.3	1293	4	US-09-170-496D-292
C 4	2902	24.3	1293	4	US-09-364-425B-57
C 5	2842	23.6	1273	4	US-09-170-496D-289
C 6	2842	23.6	1273	4	US-09-364-425B-54
C 7	2824	23.7	1279	4	US-09-170-496D-293
C 8	2824	23.7	1279	4	US-09-364-425B-58
C 9	2819	23.4	1286	4	US-09-170-496D-291
C 10	2819	23.4	1286	4	US-09-364-425B-56
C 11	2773.5	23.3	1284	4	US-09-170-496D-294
C 12	2773.5	23.3	1284	4	US-09-364-425B-59

13	2449	20.5	1056	4	US-09-513-783A-32	Sequence 32, Appl
14	2449	20.5	1056	4	US-09-430-656-32	Sequence 32, Appl
15	2384.5	20.0	541	4	US-09-513-783A-34	Sequence 34, Appl
16	2384.5	20.0	541	4	US-09-430-656-34	Sequence 34, Appl
17	2360.5	19.8	642	2	US-08-818-253-6	Sequence 6, Appl
18	2360.5	19.8	642	3	US-08-818-253-6	Sequence 6, Appl
19	2360.5	19.8	652	2	US-08-818-253-4	Sequence 4, Appl
20	2360.5	19.8	652	3	US-08-818-253-4	Sequence 4, Appl
21	2359.5	19.8	642	2	US-08-818-253-2	Sequence 2, Appl
22	2359.5	19.8	642	3	US-08-818-253-2	Sequence 2, Appl
23	2337.5	19.6	656	2	US-08-818-253-8	Sequence 8, Appl
24	2337.5	19.6	656	3	US-08-818-253-8	Sequence 8, Appl
25	2216.5	18.6	1610	4	US-09-513-783A-22	Sequence 22, Appl
26	2216.5	18.6	1610	4	US-09-430-656-22	Sequence 22, Appl
27	1774	14.9	360	5	PCT-US93-01676A-2	Sequence 4, Appl
28	1774	14.9	370	5	PCT-US93-01676A-4	Sequence 4, Appl
29	1768	14.8	360	3	US-09-025-769B-293	Sequence 293, App
30	1768	14.8	360	4	US-09-490-070A-293	Sequence 293, App
31	1768	14.8	360	4	US-09-490-153-293	Sequence 293, App
32	1768	14.8	360	4	US-09-490-324-293	Sequence 293, App
33	1765	14.8	1049	3	US-09-522-666-6	Sequence 6, Appl
34	1759	14.8	441	3	US-09-522-666-4	Sequence 4, Appl
35	1759	14.8	1024	3	US-09-522-666-2	Sequence 2, Appl
36	1754	14.7	360	1	US-08-456-923-21	Sequence 21, Appl
37	1569	13.2	319	3	US-09-044-796A-13	Sequence 13, Appl
38	1569	13.2	319	4	US-09-725-460A-13	Sequence 13, Appl
39	1510	12.7	353	4	US-09-489-039A-8090	Sequence 8090, Ap
40	1503	12.6	299	3	US-09-025-769B-285	Sequence 285, App
41	1503	12.6	299	3	US-09-025-769B-298	Sequence 298, App
42	1503	12.6	299	3	US-09-025-769B-300	Sequence 300, App
43	1503	12.6	299	4	US-09-490-070A-285	Sequence 285, App
44	1503	12.6	299	4	US-09-490-070A-298	Sequence 298, App
45	1503	12.6	299	4	US-09-490-070A-300	Sequence 300, App

ALIGNMENTS

RESULT 1
US-09-170-496D-290
; Sequence 290, Application US/09170496D
; Patent No. 6555339
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/09/170,496D
; CURRENT FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 290
; LENGTH: 1310
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-170-496D-290

Alignment Scores:
Pred. No.: 2.15e-238 Length: 1310
Score: 3004.00 Matches: 694
Percent Similarity: 65.51% Conservative: 45
Best Local Similarity: 61.52% Mismatches: 165
Query Match: 24.91% Indels: 224
DB: 4 Gaps: 42

US-10-649-433C-1 (1-6729) x US-09-170-496D-290 (1-1310)

QY 6352 TCTGATCTTGGCAACGACATCGCAGTGGGACGATCCCTCATTACGATTGCTGCG 6293
:::|||||:::|||||
Db 84 ThrAspGluIrrpGlu-----GlnTrp-----trp 91
QY 6292 -----TTTGTGTAACCGGACATCGCAGTCCGCTTCGCTCGCTATCGGCT 6239

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: August 2, 2005, 18:23:19 ; Search time 801 Seconds
(without alignments)
6545.998 Million cell updates/sec

Title: US-10-649-433C-1
Perfect score: 11920
Sequence: 1 gttgacagcttatcatcg.....agttagcggaattgatctg 6729

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0
Searched: 1745140 seqs, 389608008 residues

Total number of hits satisfying chosen parameters: 3490280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2_1/USPTO.spool_p/US10649433/runat_02082005_164946_1571/app_query.fasta_1.6919
-DB=Published Applications AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -LOCAL=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10649433 @CGN 1 1 964 @runat 02082005 164946_1571
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSFBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DSEXT=7

Database :
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8639.5	72.5	1967	16	US-10-477-044-1 Sequence 1, Appli
c 2	3004	24.9	1310	14	US-10-251-385-290 Sequence 290, App
c 3	3004	24.9	1310	18	US-10-668-035-55 Sequence 55, Appl
4	2902	24.3	1293	14	US-10-251-385-292 Sequence 292, App
5	2902	24.3	1293	18	US-10-668-035-57 Sequence 57, Appl
c 6	2842	23.6	1273	14	US-10-251-385-289 Sequence 289, App
c 7	2842	23.6	1273	18	US-10-668-035-54 Sequence 54, Appl
8	2824	23.7	1279	14	US-10-251-385-293 Sequence 293, App
9	2824	23.7	1279	18	US-10-668-035-58 Sequence 58, Appl
c 10	2819	23.4	1286	14	US-10-251-385-291 Sequence 291, App
c 11	2819	23.4	1286	18	US-10-668-035-56 Sequence 56, Appl
12	2773.5	23.3	1284	14	US-10-251-385-294 Sequence 294, App
13	2773.5	23.3	1284	18	US-10-668-035-59 Sequence 59, Appl
14	2486.5	20.9	740	15	US-10-344-404-12 Sequence 12, Appl
15	2486.5	20.9	740	15	US-10-344-404-27 Sequence 27, Appl
16	2486.5	20.9	764	15	US-10-344-404-34 Sequence 34, Appl
17	2463.5	20.7	740	15	US-10-344-404-23 Sequence 23, Appl
18	2449	20.5	1056	14	US-10-100-957A-32 Sequence 32, Appl
19	2416.5	20.3	734	15	US-10-344-404-31 Sequence 31, Appl
20	2416.5	20.3	758	15	US-10-344-404-36 Sequence 36, Appl
21	2413	20.2	745	15	US-10-344-404-19 Sequence 19, Appl
22	2395	20.1	823	15	US-10-344-404-42 Sequence 42, Appl
23	2387.5	20.0	762	15	US-10-344-404-40 Sequence 40, Appl
24	2386.5	20.0	763	15	US-10-344-404-38 Sequence 38, Appl
25	2384.5	20.0	541	14	US-10-100-957A-34 Sequence 34, Appl
26	2371	19.9	723	17	US-10-732-923-14279 Sequence 14279, A
27	2360.5	19.8	642	9	US-09-554-000-6 Sequence 6, Appli
28	2360.5	19.8	652	9	US-09-554-000-4 Sequence 4, Appli
29	2359.5	19.8	642	9	US-09-554-000-2 Sequence 2, Appli
30	2359	19.8	803	16	US-10-634-740-7 Sequence 7, Appli
31	2350	19.7	635	16	US-10-634-740-5 Sequence 5, Appli
32	2337.5	19.6	656	9	US-09-554-000-8 Sequence 8, Appli
33	2216.5	18.6	1610	14	US-10-100-957A-22 Sequence 22, Appl
34	2201.5	18.5	607	10	US-09-293-670-55 Sequence 55, Appl
35	2192.5	18.4	477	10	US-09-293-670-54 Sequence 54, Appl
36	1788	15.0	363	10	US-09-987-763-29 Sequence 29, Appl
37	1771	14.9	372	16	US-10-469-881-3 Sequence 3, Appli
38	1765	14.8	360	13	US-10-197-053-1 Sequence 1, Appli
39	1758	14.7	360	13	US-10-197-053-2 Sequence 2, Appli
40	1579	13.2	332	13	US-10-197-053-6 Sequence 6, Appli
41	1579	13.2	332	17	US-10-893-671-82 Sequence 82, Appl
42	1578	13.2	309	16	US-10-840-796-1 Sequence 1, Appli
43	1569	13.2	319	9	US-09-725-460A-13 Sequence 13, Appl
44	1569	13.2	319	20	US-11-004-853-15 Sequence 15, Appl
45	1486	12.5	1213	15	US-10-381-882-4 Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-10-477-044-1
; Sequence 1, Application US/10477044
; Publication No. US20040265902A1
; GENERAL INFORMATION:
; APPLICANT: Frickner, Mark
; APPLICANT: Vaux, David
; TITLE OF INVENTION: Universal Fluorescent Sensors
; FILE REFERENCE: HO-P02985US0
; CURRENT APPLICATION NUMBER: US/10/477,044
; CURRENT FILING DATE: 2003-11-07
; PRIOR APPLICATION NUMBER: PCT/GB02/02183
; PRIOR FILING DATE: 2002-02-10
; PRIOR APPLICATION NUMBER: GB 0111459.4
; PRIOR FILING DATE: 2001-05-10
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1967
; TYPE: PRT

ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: PLASMIID pTrocFRET3
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: [540]..(1949)
OTHER INFORMATION: X=unknown
US-10-477-044-1

Alignment Scores:
Pred. No.: 0 Length: 1967
Score: 8639.50 Matches: 1701
Percent Similarity: 81.30% Conservative: 21
Best Local Similarity: 80.31% Mismatches: 76
Query Match: 72.48% Indels: 320
DB: 16 Gaps: 8

US-10-649-433C-1 (1-6729) x US-10-477-044-1 (1-1967)

```
QY 413 ATGTTGAGCAAGGCGAGAGCTGTTCAACCGGGTGGTCCCATCTCTGGTCGAGCTGGAC 472
Db 4 MetValSerLysGlyGluLeuPheThrGlyValValProIleLeuValGluLeuAsp 23
QY 473 GCGCAGCTAAACGCGCACAGTTACGCTGTCCGCGAGGGCGAGGGCGATGCCACCTAC 532
Db 24 GlyAspValAsnGlyHisLysPheSerValSerGlyGluGlyGluGlyAspAlaThrTyr 43
QY 533 GCGCAGCTGACCTCAAGTTTCATCTGCACACCGGCAAGCTGCCGTGCCCTGGCCAC 592
Db 44 GlyLysLeuThrLeuLysPheIleCysThrThrGlyLysLeuProValProTyrProThr 63
QY 593 CTCGTGACACCTTCGGCTACGGCTGACGCTGCTGCGCTGCTGCGCTGCTGCGCTGCTG 652
Db 64 LeuValThrThrLeuThrTyrGlyValGlnCysPheSerArgTyrProAspHisMetLys 83
QY 653 CAGCAGCACTCTTCAAGTCCGCCATGCCGAGGCTACGCTCCAGAGGCGCACCATCTTC 712
Db 84 GlnHisAspPheLysPheLysAlaMetProGluGlyTyrValGlnGluArgThrIlePhe 103
QY 713 TTCAGGACGACGCGCAACTACAAAGCCGCGCGAGGTGAAGTTCGAGGGCGACACCTG 772
Db 104 PheLysAspGlyAsnTyrLysThrArgAlaGluValLysPheGluGlyAspThrLeu 123
QY 773 GTGAACCGCATCGACTGAAGGCGATCGACTTCAAGGAGGCGGCAACATCTCTGGGCGAC 832
Db 124 ValAsnArgIleGluLeuLysGlyIleAspPheLysGluAspGlyAsnIleLeuGlyHis 143
QY 893 AAGCTGGAGTACAACTACAAAGCCACCAACGCTCTATCATCGCGCGCACAGCAGAGAAC 892
Db 144 LysLeuGluTyrAsnTyrIleSerHisAsnValTyrIleThrAlaAspLysGlnLysAsn 163
QY 893 GGCATCAAGGTGAATCTCAAGATCCGCGCACCAACATCGAGGAGGCGGCGAGCTCGCC 952
Db 164 GlyIleLysAlaAsnPheLysIleArgHisAsnIleGluAspGlySerValGlnLeuAla 183
QY 953 GACCATACAGCAGAAACCCCGCATCGCGGAGGCGGCGGCGGCTGCTCTGCGCGCAAC 1012
Db 184 AspHisTyrGlnGlnAsnThrProIleGlyAspGlyProValLeuLeuProAspAsnHis 203
QY 1013 TACCTGAGTACACGCTCGCGCTGAGCAAGAACCCCAAGAGCGGATCACATGTC 1072
Db 204 TyrLeuSerThrGlnSerAlaLeuSerLysAspProAsnGluLysArgAspHisMetVal 223
QY 1073 CTGCTGGAGTTCGTACCGCGCC-----GGGATCACTCTC 1108
Db 224 LeuLeuGluPheValThrAlaAlaLeuGlnSerSerGlyGlyGlyGlySerThrMet 243
QY 1109 GCGATGGACGAG-----CTGTACAGACTAGTCTGCTACTCGCAT 1150
Db 244 GlyGlySerHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 263
QY 1151 GGTGTAAACAATCTATAAGTACCAACATACTTTATGCTGTAGTGGCAAGGCTATTGAG 1210
Db 1151 GGTGTAAACAATCTATAAGTACCAACATACTTTATGCTGTAGTGGCAAGGCTATTGAG 1210
```

```
Db 264 GlyArgAspLeuTyrAspAspAsp----- 272
QY 1211 CAAGATCGAAAGCGCGCCAGATGTTCAAGCTGCTGATGAATGATTTCTCAGATGACCCAG 1270
Db 272 ----- 272
QY 1271 TCCAAAGCAGACGATCAGATCGACGTATTGCTGGCCAAAGGGGTGAAGGCACCTGGCCATC 1330
Db 272 ----- 272
QY 1331 AACCTGGTTGACCGCGCAGCTCGCGGTACGGGTACGGTATGAGAAAGCGCGTGGCAAAACGTG 1390
Db 272 ----- 272
QY 1391 CCGGTGTTTTTCTTCAAAAGAAACCGCTCTCGTAAAGGCGTGGATAGCTACGACAAAGCC 1450
Db 272 ----- 272
QY 1451 TACTACGTTGCACTGACTCAAAAGAGTCCGGCATTTTCAAGGCGATTTGATTCCTAAA 1510
Db 273 ----- 273
QY 1511 CACTGGCGCGCAATCAGGGTTGGATCTGAACAAAGACGGTCAGATTCACTGTTCTACTG 1570
Db 274 His-----ArgTyrIleArg 278
QY 1571 CTGAAAGGTGAACCGGGCCATCCGGATGCAAGACGATACCTACCTTACGTTGATTAAGAA 1630
Db 279 ProArgGlySerSerGly----- 284
QY 1631 TTGAACGATAAAGCATCAAAACTGAACAGTTACAGTTAGATACCGCAATGTGGGACACC 1690
Db 284 ----- 284
QY 1691 GCTCAGCGAAAGATAAGATGAGCGCCCTGGCTCTGGCCCGAAGCCCAAAATCGAA 1750
Db 284 ----- 284
QY 1751 GTGGTTATCGCAACAACGATGCGATGGCAATGGCGGGTGAAGCGCTGAAGACACAC 1810
Db 284 ----- 284
QY 1811 AACAACTCCAGCATTCGGGTGTTGGCGTCGATGCGTCCAGAAAGCGCTGGCGTGTG 1870
Db 284 ----- 284
QY 1871 AAATCCGTCGATCGCGGCGCACCGTACTGAACGATGCTAAACAACGCGGAAAGCGACC 1930
Db 285 -----GlyGlyGlySerGly----- 289
QY 1931 TTTGATCTGGCGAAAAACCTCGCGCGATGTTAAAGTGGCGGTGATGGCACCAACTGGAAA 1990
Db 290 -----GlyGlyGlySerGlyGly----- 296
QY 1991 ATCGACAACAAGTGGTCCGCGTACCTTATGTTGGCGTAGATAAAGACACACCTGGCTGAA 2050
Db 296 ----- 296
QY 2051 TTCAGCAAGAAAGGTACCAAGTAAAGGAGAAACCTTTTCACTGGAGTTGTCCCAATCTT 2110
Db 297 ---SerSerArgMetValSerLysGlyGluLeuPheThrGlyValValProIleLeu 315
QY 2111 GTTGAATTAGATGGTGAATTAATGGGCACAAATTTTCTGTCAGTGGAGAGGTGAAGGT 2170
Db 316 ValGluLeuAspGlyAspValAsnGlyHisLysPheSerValSerGlyGluGlyGly 335
QY 2171 GATGCAACATACGAAACCTTACCTTAAATTTTTCACCTACTCTGCACTGCACTGCTGTT 2230
Db 336 AspAlaThrTyrGlyLysLeuThrLeuLysPheIleCysThrThrGlyLysLeuProVal 355
QY 2231 CCATGGCCAAACCTGTCTACTTCTTATGTTGTTCAATGTTTCCGTTATCCG 2290
Db 356 ProTyrProThrLeuValThrThrPheGlyTyrGlyLeuGlnCysPheAlaArgTyrPro 375
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 4, 2005, 19:37:23 ; Search time 27846 Seconds
(without alignments)
11709.227 Million cell updates/sec

Title: US-10-649-433C-1

Perfect score: 6729

Sequence: 1 gttgacagcttatcatcga.....agttagcgcgaattgatctg 6729

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_scs:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4344.4	64.6	5271	12	CVU89936
2	3855	57.3	4806	12	AF050464
3	3854.2	57.3	4205	6	AX576406
4	3853.4	57.3	5013	6	AX264513
5	3853.2	57.3	4203	6	AX576404
6	3853.2	57.3	4204	6	AX576405
7	3853.2	57.3	4217	6	AX576408
8	3853.2	57.3	4218	6	AX576409
9	3852.8	57.3	4216	6	AX576407
10	3852.2	57.2	4176	6	AI3038
11	3852.2	57.2	4176	12	SYNTRC99A
12	3852.2	57.2	4239	6	A29289
13	3851.6	57.2	4476	6	ARI02989
14	3851.6	57.2	4476	6	ARI172783
15	3851.4	57.2	4214	6	BD273391
16	3851.4	57.2	4214	6	AX027488
17	3850.6	57.2	4176	12	XXU13872
18	3824.8	56.8	4232	12	SYNLACIQ
19	3824.8	56.8	4735	12	SYNBETALAC

c	20	3391.8	50.4	6862	6	AI0361
c	21	3382.4	50.3	4733	12	XXU19585
c	22	3375.2	50.2	4361	12	AY327137
c	23	3375.2	50.2	5189	12	AY327139
c	24	3344.8	49.7	5024	6	CQ846977
c	25	3343.2	49.7	4803	12	AY243506
c	26	3341.2	49.7	6464	6	BD263349
c	27	3304.4	49.1	4770	12	AF333022
c	28	3297.4	49.0	5885	1	STAFRA
c	29	3230.2	48.0	5398	6	AX824326
c	30	2889.8	42.9	8811	12	AY569779
c	31	2889.8	42.9	12404	12	AY569780
c	32	2864.8	42.6	4109	12	PSPORT1
c	33	2864.8	42.6	5957	6	BD263353
c	34	2864.8	42.6	5957	6	BD263354
c	35	2863.2	42.6	5451	6	AX771237
c	36	2859.6	42.5	4932	12	SYNECOPGEX
c	37	2859.6	42.5	4968	12	SYNGLUPTAN
c	38	2859.6	42.5	4983	12	CVU78874
c	39	2859.6	42.5	4984	12	CVU78872
c	40	2859.6	42.5	4985	12	CVU78873
c	41	2859.6	42.5	5487	12	FGH223813
c	42	2858	42.5	4506	12	AB116367
c	43	2858	42.5	4557	6	EI3740
c	44	2858	42.5	4557	6	AR373251
c	45	2858	42.5	4947	12	XXU13849

ALIGNMENTS

RESULT 1

CVU89936

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CVU89936 Cloning vector p6xHis-GFP, complete sequence.

U89936 5271 bp DNA circular SYN 01-AUG-1997

U89936.1 GI:2190721

unidentified cloning vector

unidentified cloning vector

other sequences; artificial sequences; vectors.

1 (bases 1 to 5271)

Kitts, P.

p6xHis-GFP, complete sequence

Unpublished

(Bases 1 to 5271)

Holtz, A. and Lou, Y.

Direct Submission

Submitted (18-FEB-1997) CLONTECH Laboratories, Inc., 1020 East

Meadow Circle, Palo Alto, CA 94303-4230, USA

This vector can be obtained from CLONTECH Laboratories, Inc., 1020

East Meadow Circle, Palo Alto, CA 94303-4230, USA. To place an

order, call (415) 424-8222 or (800) 662-2566, extension 1.

International customers, please contact your local distributor or

subsidiary. For technical information, call (415) 424-8222 or (800)

662-2566, extension 3.

This sequence has been compiled from information in the sequence

databases, published literature and other sources, together with

partial sequences obtained by CLONTECH. If you suspect there is an

error in this sequence, please contact CLONTECH's Technical Support

Department at (415) 424-8222 or (800) 662-2566, extension 3 or

E-mail TECH@CLONTECH.COM.

Location/Qualifiers

1..5271

/organism="unidentified cloning vector"

/mol_type="genomic DNA"

/db_xref="taxon:45196"

ORIGIN

Query Match 64.6%; Score 4344.4; DB 12; Length 5271;

Best Local Similarity 95.0%; Pred. No. 0;

Matches 4582; Conservative 0; Mismatches 81; Indels 159; Gaps 3;

QY 2067 CCAGTAAAGGAGAACTTTTCACTCGAGTTGTCCCAATCTCTGTTGAAATTAGATGGT 2126
DB |||||
DB 450 CTAGCAAAAGGAGAACTTTTCACTGGAGTTGTCCCAATCTCTGTTGAAATTAGATGGT 509
QY 2127 ATGTTAAATGGGCACAAATTTTCTGTCACTGGAGAGGGTGAAGGTGATGCAACATACGGAA 2186
DB |||||
DB 510 ATGTTAAATGGGCACAAATTTTCTGTCACTGGAGAGGGTGAAGGTGATGCAACATACGGAA 569
QY 2187 AACTTACCTTAAATTTATTTGCACTACTGGAATACTACTGTTCCATGGGCCAACACTTG 2246
DB |||||
DB 570 AACTTACCTTAAATTTATTTGCACTACTGGAATACTACTGTTCCATGGGCCAACACTTG 629
QY 2247 TCACCTACTTCTTATGGTGTCAATGTTTCCGGTTATCCGGATCATATGAAACGGC 2306
DB |||||
DB 630 TCACCTACTTCTTATGGTGTCAATGTTTCAAGATACCCAGATCATATGAAACGGC 689
QY 2307 ATGACTTTTCAAGAGTGCATGCCGGAAGGTTATGTACAGGAACGCACTATCTTTCA 2366
DB |||||
DB 690 ATGACTTTTCAAGAGTGCATGCCGGAAGGTTATGTACAGGAACGCACTATCTTTCA 749
QY 2367 AAGATGCGGGAACCTACAAGACGCTGTGAAGTCAAGTTTGAAGGTGATACCCCTTGTTA 2426
DB |||||
DB 750 AAGATGCGGGAACCTACAAGACACGCTGTGAAGTCAAGTTTGAAGGTGATACCCCTTGTTA 809
QY 2427 ATCGTATCGAGTTAAAGGTATTGATTTTAAAGAGATGGAACATTTCTCGACACAAAC 2486
DB |||||
DB 810 ATAGAATCGAGTTAAAGGTATTGATTTTAAAGAGATGGAACATTTCTTGACACAAAT 869
QY 2487 TCGAGTCAAACTATAAATCACACAATGTATACATCAGCGACACAAACAAAGAAATGGAA 2546
DB |||||
DB 870 TGGATACAACCTATAAATCACACAATGTATACATCAGCGACACAAACAAAGAAATGGAA 929
QY 2547 TCAAAAGTAACTTCAAAATTCGCCACAACTTGAAGATGGATCGGTTCAACTAGCAGACC 2606
DB |||||
DB 930 TCAAAAGTAACTTCAAAATTAGACACAACTTGAAGATGGAAGCTTCAACTAGCAGACC 989
QY 2607 ATTATCAACAAATCTCCAAATGGCGATGGCCCTGCTCTTTTACGACACCAATTACC 2666
DB |||||
DB 990 ATTATCAACAAATCTCCAAATGGCGATGGCCCTGCTCTTTTACGACACCAATTACC 1049
QY 2667 TGTGACACAACTCTGCCCTTTCGAAAGATCCCAACGAAAGCGTGACACACATGTCCTTC 2726
DB |||||
DB 1050 TGTCCACACAACTCTGCCCTTTCGAAAGATCCCAACGAAAGAGACACACATGTCCTTC 1109
QY 2727 TTGAGTTTGAATCTGCTGGGATTTACATGGCATGGATGGATGCTCTACAAATAAAG- 2785
DB |||||
DB 1110 TTGAGTTTGAATCTGCTGGGATTTACATGGCATGGATGGAATATACAAATAAATGT 1169
QY 2786 -----CTTACGTAGAACAAACTCA----- 2806
DB |||||
DB 1170 CCAGACTTCCAAATTGACACTAAGTGTCCGAACAATTTACTAANAATCTCAGGGTTCCGTGT 1229
QY 2807 ----- 2806
DB 1230 TAAATTCAGCTGAGATATTATTTATATATTTATAGATTCATTAATACTGTATGAATAAT 1289
QY 2807 -----TCTCAGAAAGAGATCTGAATAGCGCCGTCG 2836
DB |||||
DB 1290 TTAATTGATGTTAATTGATAGAGTTATTTTCTTAATTAACACAGGCTACTTGGAGTGTATTCT 1349
QY 2837 ACCATCATCATCATCATGATGAGTTTAAACGGTCTC----- 2873
DB |||||
DB 1350 TAAATCTATATTAATTACAATTTGAATTTGACTTTCAAAAAGCGCGCGCGCGCGCG 1409
QY 2874 -----CAGCTTGGCTGTTTGGCGGATGAGAAAGATTTTTCAGCCCTGATACAGATTAAA 2927
DB |||||
DB 1410 AATTCGAAGCTTGGCTGTTTGGCGGATGAGAAAGATTTTTCAGCCCTGATACAGATTAAA 1469
QY 2928 TCAGAACCGAGAGCGGTCTGATAAAACAGAAATTTGCCCTGGCGGAGTAGCGCGTGTCT 2987
DB |||||
DB 1470 TCAGAACCGAGAGCGGTCTGATAAAACAGAAATTTGCCCTGGCGGAGTAGCGCGTGTCT 1529
QY 2988 CCACCTGACCCCATGCCGAATCTCAGAAAGTGAACCGCGTAGCGCGGATGTGTGGG 3047

DB 1530 CCACCTGACCCCATGCCGAATCTAGAAAGTGAACCGCGTAGCGCGATGTTGTGTGGG 1589
QY 3048 TCTCCCATCGAGAGTAGGAACTGCGAGGCATCAATAAAGCAAAAGGCTCAGTTCGAA 3107
DB |||||
DB 1590 TCTCCCATCGAGAGTAGGAACTGCGAGGCATCAATAAAGCAAAAGGCTCAGTTCGAA 1649
QY 3108 AGACTGGGCTTTTCGTTTATCTGTTTGTCTGGTGAACGCTCTCTCTGAGTAGGACAAA 3167
DB |||||
DB 1650 AGACTGGGCTTTTCGTTTATCTGTTTGTCTGGTGAACGCTCTCTCTGAGTAGGACAAA 1709
QY 3168 TCCGCGGAGCGGATTTGAACGTTTGAAGCAACCGCGAGGCTGCGGGCAGAGACG 3227
DB |||||
DB 1710 TCCGCGGAGCGGATTTGAACGTTTGAAGCAACCGCGAGGCTGCGGGCAGAGACG 1769
QY 3228 CCGGCCATAAATCCAGGCATCAAAATTAAGCAGAAGCCATCTCTGACGAGTGCCTTTT 3287
DB |||||
DB 1770 CCGGCCATAAATCCAGGCATCAAAATTAAGCAGAAGCCATCTCTGACGAGTGCCTTTT 1829
QY 3288 TCGGTTTCTACAACTCTTTTGTATTCTTAATAACATTCAAATATGATATCCGCTC 3347
DB |||||
DB 1830 TCGGTTTCTACAACTCTTTTGTATTCTTAATAACATTCAAATATGATATCCGCTC 1889
QY 3348 ATGAGACAAATAACCTCGATAAATCTCAATAATATTGAAAGAGAGATGAGTATT 3407
DB |||||
DB 1890 ATGAGACAAATAACCTCGATAAATCTCAATAATATTGAAAGAGAGATGAGTATT 1949
QY 3408 CAACATTTCCGTGTCGCCCTTATTTCCCTTTTTCGGCATTTTGCCTTCTCTGTTTTCCT 3467
DB |||||
DB 1950 CAACATTTCCGTGTCGCCCTTATTTCCCTTTTTCGGCATTTTGCCTTCTCTGTTTTCCT 2009
QY 3468 CACCAGAAACGCTGGTGAAGATGCTGAAGATCAAGTTGGGTGACGAGTGGGT 3527
DB |||||
DB 2010 CACCAGAAACGCTGGTGAAGATGCTGAAGATCAAGTTGGGTGACGAGTGGGT 2069
QY 3528 TACATCGAACTGGATCTCAACAGCGTAAGATCTTGAAGAGTTTTCGCCCGCAAGAAAGCT 3587
DB |||||
DB 2070 TACATCGAACTGGATCTCAACAGCGTAAGATCTTGAAGAGTTTTCGCCCGCAAGAAAGCT 2129
QY 3588 TTTCCAAATGATGACACCTTTTAAAGTTCTGTGTGGCGGTAATTAATCCGCTGTTGAC 3647
DB |||||
DB 2130 TTTCCAAATGATGACACCTTTTAAAGTTCTGTGTGGCGGTAATTAATCCGCTGTTGAC 2189
QY 3648 GCCGGCAAGAGCAACTCGGTGCGGCATACACTATTCTCAGAAATGACTTGGTTCAGTAC 3707
DB |||||
DB 2190 GCCGGCAAGAGCAACTCGGTGCGGCATACACTATTCTCAGAAATGACTTGGTTCAGTAC 2249
QY 3708 TCACCAAGTCAAGAAAGCATCTTACGGATGGCATGACAGTAAGAGAAATTAATGCAAGTCT 3767
DB |||||
DB 2250 TCACCAAGTCAAGAAAGCATCTTACGGATGGCATGACAGTAAGAGAAATTAATGCAAGTCT 2309
QY 3768 GCCATAACCATGAGTGTATAACCTGCGGCCAACTTATCTTCTGACAAACGATCGGAGACCG 3827
DB |||||
DB 2310 GCCATAACCATGAGTGTATAACCTGCGGCCAACTTATCTTCTGACAAACGATCGGAGACCG 2369
QY 3828 AAGGAGCTAAACGCTTTTTCGACAAATGGGGGATCATGTAACTCGCCTTGTATGCTGG 3887
DB |||||
DB 2370 AAGGAGCTAAACGCTTTTTCGACAAATGGGGGATCATGTAACTCGCCTTGTATGCTGG 2429
QY 3888 GAAACCGAGCTGAATTAAGCCATCAAAACGACGAGCGGTGACACACCATGCTGTAGCA 3947
DB |||||
DB 2430 GAAACCGAGCTGAATTAAGCCATCAAAACGACGAGCGGTGACACACCATGCTGTAGCA 2489
QY 3948 ATGGCAACAAAGTTGGCGCAAACTAATTAACCTGGCGAACTACTTACTCTAGCTTCCCGGAA 4007
DB |||||
DB 2490 ATGGCAACAAAGTTGGCGCAAACTAATTAACCTGGCGAACTACTTACTCTAGCTTCCCGGAA 2549
QY 4008 CAATTAATAGACTCGAGCGGATGAAGTTGAGGAGCCACTTCTCGGCTCGGCCCTT 4067
DB |||||
DB 2550 CAATTAATAGACTCGAGCGGATGAAGTTGAGGAGCCACTTCTCGGCTCGGCCCTT 2609
QY 4068 CCGGCTGCTGTTTATTTGCTGATAAAATCTGGAGCGGTGAGCGGTCTCGGCTGATC 4127
DB |||||

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 4, 2005, 19:20:28 ; Search time 3192 Seconds
(without alignments)
12479.304 Million cell updates/sec

Title: US-10-649-433C-1

Perfect score: 6729

Sequence: 1 gtttagacgttatcatcga.....agtttagcggaattgatctg 6729

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_16Dec04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001s:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002s:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003s:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004s:*
- 13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4119.8	61.2	5903	8	ABZ233939
2	3855	57.3	5027	2	AAX78865
3	3855	57.3	5069	2	AAX78871
4	3854.2	57.3	4205	6	ABQ74927
5	3853.4	57.3	4816	4	ABQ7799
6	3853.4	57.3	4816	12	ADf72546
7	3853.4	57.3	4816	12	ADG14708
8	3853.4	57.3	5013	9	ACA63355
9	3853.4	57.3	5013	10	AAD59423
10	3853.4	57.3	5013	10	ADe86127
11	3853.4	57.3	5104	4	AAD07798
12	3853.4	57.3	5104	12	ADf72545
13	3853.2	57.3	4203	6	ABQ74925
14	3853.2	57.3	4204	6	ABQ74926
15	3853.2	57.3	4217	6	ABQ74929
16	3853.2	57.3	4218	6	ABQ74930
17	3852.8	57.3	4216	6	ABQ74928
18	3852.2	57.2	5926	2	AAV32978
19	3852.2	57.2	5926	2	AAV32977
20	3852.2	57.2	5926	6	AAD45060

C	21	3852.2	57.2	5926	6	AAD45059
C	22	3852.2	57.2	5926	12	ADG46817
	23	3851.8	57.2	5069	2	AAX78874
	24	3851.8	57.2	5099	2	AAX78896
	25	3851.8	57.2	5132	2	AAX78883
	26	3851.8	57.2	5324	2	AAX78890
	27	3851.8	57.2	5342	2	AAX78893
	28	3851.4	57.2	4214	3	AAX59260
	29	3851.4	57.2	4214	3	AAX59260
	30	3850.6	57.2	4176	1	AAAG2632
	31	3850.6	57.2	4176	1	AAAG2632
C	32	3850.6	57.2	5926	12	ADG46818
	33	3841.4	57.1	5103	12	ADG14707
	34	3727.8	55.4	4189	2	AAQ05397
	35	3542	52.6	5161	10	ADH45226
C	36	3344.8	49.7	5024	9	ACF06051
C	37	3344.8	49.7	5024	12	ADL18668
	38	3341.2	49.7	6464	3	AAC55454
	39	3341.2	49.7	6464	10	ABZ58765
C	40	3241.8	48.2	7174	1	AAAG1062
	41	3230.2	48.0	5398	10	ADP50156
C	42	2864.8	42.6	5957	3	AAC55467
	43	2864.8	42.6	5957	3	AAC55464
	44	2863.2	42.6	5451	9	ADA41729
	45	2858	42.5	4557	2	AAT90491

ALIGNMENTS

RESULT 1

ABZ233939

ID ABZ233939 standard; DNA; 5903 BP.

XX AC ABZ233939;

XX DT 18-MAR-2003 (first entry)

XX DE Nucleotide sequence of plasmid pTrcCFRET3.

XX KW Fluorescent; fluorescence resonance energy transfer; FRET; tumour;

XX KW estrogen; tamoxifen; therapy; pTrcCFRET3; ds.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT CDS 2..5902

FT /*tag= a

FT /note= "contains internal stop codons"

XX PN WO200290987-A2.

XX PD 14-NOV-2002.

XX PR 10-MAY-2002; 2002WO-GB002183.

XX PR 10-MAY-2001; 2001GB-00011459.

XX PA (ISIS-) ISIS INNOVATION LTD.

XX PI Adg14708 Globalar

XX PI Aca63355 E. coli D

XX XX AAd59423 E. coli pu

XX DR ADe86127 E. coli p

XX DR AAd07798 pTrcHisB

XX PT AAdf72545 Mouse Adi

XX PT AAdg74925 Avidity I

XX PT AAdg74926 Avidity I

XX PT AAdg74929 Avidity I

XX PT AAdg74930 Avidity I

XX PT AAdg74928 Avidity I

XX PS AAv32978 Tn7 donor

XX PS AAv32977 Tn7 donor

XX CC AAd45060 pEM delta

XX CC AAd45060 pEM delta

XX CC AAd45060 pEM delta

XX CC AAd45060 pEM delta

XX CC AAd45060 pEM delta

XX CC AAd45060 pEM delta

XX CC AAd45060 pEM delta

XX CC AAd45060 pEM delta

XX CC AAd45060 pEM delta

XX CC AAd45060 pEM delta

XX CC AAd45060 pEM delta

Probe, useful e.g. for medical diagnosis, and detection of pollutants in water systems and contaminants in foodstuffs, has target binding site moiety and mimic moiety attached to two fluorescent polypeptides, and linker.

Disclosure; Fig 5A-H; 50pp; English.

The invention relates to a probe (I) comprising a target binding site

CC moiety attached to a first fluorescent polypeptide, a mimic moiety capable of binding to the target binding site moiety and attached to a second fluorescent polypeptide, and a linker connecting the two fluorescent polypeptides and which allows the distance between the fluorescent polypeptides to vary. (I), or a cell harbouring (I), or a polynucleotide which encodes (I), or a sensor (V) comprising (I), a light source which is capable of exciting the probe, and a detecting which is capable of measuring the amount of fluorescence resonance energy transfer (FRET) from the probe are useful for detecting the presence or absence of a target substance in a test sample, and for identifying an inhibitor of binding between two substances, where the two substances would bind to each other in the absence of an inhibitor. (I) is useful for medical diagnosis, detection of pollutant in water systems, and detection of contaminants in foodstuffs and in animal and plant biology and for identification of new therapeutic substances. (I) is useful to screen for stimulators, which increase or promote binding between two substances, to identify a factor which increases the strength of binding between two substances or a factor whose presence is necessary for the binding of two substances to take place, to identify stimulators and/or stabilizers of binding interactions, to detect the presence of a substance, for e.g. a metabolite, hormone, drug, toxin, or pollutant in an extract, for e.g. a fluid sample derived from any organism, including an animal or human, plant, fungus or microbe. (I) is also useful for detecting sugars, oligosaccharides or non-carbohydrate mimetics, to determine the presence or absence of steroid hormones, for diagnosis, drug and alcohol testing and testing for exposure to toxins or pollutants, to detect air-borne substances for e.g. atmospheric pollutants if these substances are soluble, and to detect specific substances in plant, fungal or microbial for e.g. bacterial extracts. Plant extracts for e.g. exudates are useful in determining the presence of plant pathogenic viruses or bacteria in plant. Additionally, the probes may be used to determine the presence and amount of trace elements or pollutants in plant extracts. Thus results of such assays may provide indirect measurements of soil quality and as indicative of particular types of soil pollution. Further application of the probes is to detect proteins expressed in transgenic plants or transgenic animals, fungi or microbes, and to detect contaminants or pollutants in water supplies, soil or factory effluents, and in quality control for detecting substances, for e.g. contaminants in foodstuffs and medicaments. The probes are also useful for screening for inhibitors of growth of estrogen-sensitive breast tumours, which is useful for identifying anti-tumour agents that act at a site distinct from that targeted by the synthetic estrogen, tamoxifen. The probe is also useful for identifying protease inhibitors, screening for active binding site inhibitors of the protease, and for identifying intracellular G protein signal inhibitors. The inhibitors identified by the probe are useful for treating human or animal body by therapy. The present sequence represents the nucleotide sequence of plasmid pTrcCFRET3

XX Sequence 5903 BP; 1429 A; 1594 C; 1608 G; 1272 T; 0 U; 0 Other;

Query Match 61.2%; Score 4119.8; DB 8; Length 5903;
Best Local Similarity 93.5%; Pred. No. 0;
Matches 4357; Conservative 0; Mismatches 192; Indels 112; Gaps 1;

Qy 2069 AGTAAGGAGAGAACTTTTCTGTCAGTGGAGGGTGAAGTATGCAACATACGAAAA 2188
Db 905 AGCAAGGGCGAGGAGCTTTCACCGGGGTGGTCCATCTCGTGGAGCGCGGAC 964
Qy 2129 GTTAATGGGCAAAATTTTCTGTCAGTGGAGGGTGAAGTATGCAACATACGAAAA 2188
Db 965 GTAAAGGGCCACAGTTCAGGCTGTCGCGAGGGCGAGGGGATGCCACCTACGGCAAG 1024
Qy 2189 CTTACCCCTTAATTTTATTTGCACTACTGGAATACTCTGTTCATGCGCCAACTTTGTC 2248
Db 1025 CTGACCCCTGAAGTTTCATCGCACACCACCGCAAGCTGCCCGTCCGCCACCCCTCGT 1084
Qy 2249 ACTACTTCTCTTATGGTGTTCATAGCTTTTCCCGTATCCGGATCATGAAACGGCAT 2308
Db 1085 ACCACCTTCGGCTACGGGCTCGAGTGTTCGCCCGCTACCCCGACCATGAAGCAGCAC 1144
Qy 2309 GACTTTTTCAAGAGTGCCATGCCGAAGGTATGTACAGGAACGCACTATCTTTCAA 2368
Db 1145 GACTTCTTCAAGTCGCCCATGCCGAAGGTACGTTCCAGGAGCGCACCATCTTCTCAAG 1204

Qy 2369 GATCACGGGAACATAAGACGCGTCTGAAGTCTCAAGTTTGAAGGTGATACCCCTTGTAAAT 2428
Db 1205 GACGCGGCAACTACAAGACCCGCGCGAGGTGAAGTTTCGAGGGCGACACACCTTGTGAAC 1264
Qy 2429 CGTATCCAGTTAAAGGTATTGATTTTAAAGAAAGTGAACATTTCTCGGACACAAATC 2488
Db 1265 CGCATCGAGCTGAAGGGCATCGACTTCAAGGAGACGGCAACATCTCTGGGCGCAAGCTG 1324
Qy 2489 GAGTACAACATAACTCACCAATGTATACATCAGCGCAGACAAACAAAGAAATGGAATC 2548
Db 1325 GAGTACAACATAACTCACCAATGTATATATCATGCGCGACAGCAGAAACGGCATC 1384
Qy 2549 AAAGCTAACTCAAAATTCGCCCAACAAATGAAGATGATCGGTTCAACTAGCAGACCAT 2608
Db 1385 AAGGTGAATTCAGATCCGCGCACCAACATCAGGACGCGCAGCTGCGACGCCAC 1444
Qy 2609 TATCAACAAATATCTCCAATTTGGGATGGCCCTCTCTTTTACCAGACCAACATACCTG 2668
Db 1445 TACCAGCAGAACACCCCATCGCGGACGGCCCGTGTCTGCTGCCGCAACCACTACCTG 1504
Qy 2669 TCGACACAATCTGCCCTTTTCGAAAGATCCCAACGAAAGCGGTGACCAATGCTCTCTT 2728
Db 1505 AGCTACCACTCCGCGCTGACGCAAGAGACCCCAACGAGAGCGCGATCATGCTCTGCTG 1564
Qy 2729 GAGTTTGTACTGTCTGTGGATTTACATGCGCATGAGTACTCTACAAATAAAAGCTT 2788
Db 1565 GAGTTCTGTGACCGCGCGGGATCACTCTCGGCA----- 1598
Qy 2789 ACGTAGAACAAAAAATCATCTCAGAACGAGGATCTGAATAGCGCGTCGACCATCATC 2848
Db 1599 ----- 1598
Qy 2849 ATCATCATTTGATTTTAAACCGTCTCCAGCTTGGCTGTTTGGCGGATGAGAGAATTTT 2908
Db 1599 -----AGCTTGGCTGTTTGGCGGATGAGAGAATTTT 1632
Qy 2909 CAGCTGATACAGATTTAAATCAGAACGAGCGGTCTGTATATAAACAAGATTTTGCCTGG 2968
Db 1633 CAGCTGATACAGATTTAAATCAGAACGAGCGGTCTGTATATAAACAAGATTTTGCCTGG 1692
Qy 2969 CGGCAGTAGCGGTGTCTCCACCTGACCCCATCCGAACTCAGAACTGAAACGCGCTAG 3028
Db 1693 CGGCAGTAGCGGTGTCTCCACCTGACCCCATCCGAACTCAGAACTGAAACGCGCTAG 1752
Qy 3029 CGGCAGTAGGTAGTGTGGGTCTCCCATGCGAGAGTAGGGAACCTGCCAGGCATCAAAATA 3088
Db 1753 CGCCGATGGTAGTGTGGGTCTCCCATGCGAGAGTAGGGAACCTGCCAGGCATCAAAATA 1812
Qy 3089 AACGAAAGGCTCAGTCGAAAGACTGGGCTTTCTGTTTATCTGTTTGTTCGTTGAACG 3148
Db 1813 AACGAAAGGCTCAGTCGAAAGACTGGGCTTTCTGTTTATCTGTTTGTTCGTTGAACG 1872
Qy 3149 CTCTCTGAGTAGGACAAATCCGCGGAGCGGATTTGAACGTTGCGAAGCAACGCGCCG 3208
Db 1873 CTCTCTGAGTAGGACAAATCCGCGGAGCGGATTTGAACGTTGCGAAGCAACGCGCCG 1932
Qy 3209 GAGGTGCGGCGGAGGACGCCCGCATATAAACTGCCAGGCATCAAAATTAAGCAGAGGCCA 3268
Db 1933 GAGGTGCGGCGGAGGACGCCCGCATATAAACTGCCAGGCATCAAAATTAAGCAGAGGCCA 1992
Qy 3269 TCCTGACGGATGGCTTTTTCGGTTTCTAAGAACTCTTTTGTGTTTATTTTCTAAATACA 3328
Db 1993 TCCTGACGGATGGCTTTTTCGGTTTCTAAGAACTCTTTTGTGTTTATTTTCTAAATACA 2052
Qy 3329 TTCAAAATATGATCCGCTCATGAGACATAACCCCTGATAAATGCTTCAATATATTGAAA 3388
Db 2053 TTCAAAATATGATCCGCTCATGAGACATAACCCCTGATAAATGCTTCAATATATTGAAA 2112
Qy 3389 AAGGAAGATGATGATTTTCAACATTTTCGGTGTCCGCTTATTCCTTTTTTGGGCAAT 3448
Db 2113 AAGGAAGATGATGATTTTCAACATTTTCGGTGTCCGCTTATTCCTTTTTTGGGCAAT 2172

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 4, 2005, 09:18:00 ; Search time 992 Seconds
(without alignments)
11099.300 Million cell updates/sec

Title: US-10-649-433C-1

Perfect score: 6729

Sequence: 1 gtttagacgcttatcatcga.....agttagcgcgaattgatctg 6729

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2_6/prodata/1/ina/5A-COMB.seq.*
- 2: /cgn2_6/prodata/1/ina/5B-COMB.seq.*
- 3: /cgn2_6/prodata/1/ina/6A-COMB.seq.*
- 4: /cgn2_6/prodata/1/ina/6B-COMB.seq.*
- 5: /cgn2_6/prodata/1/ina/PCTUS-COMB.seq.*
- 6: /cgn2_6/prodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3853.4	57.3	5013	US-09-183-188B-5	Sequence 5, Appli
2	3852.2	57.2	5226	US-09-027-169-3	Sequence 3, Appli
3	3852.2	57.2	5226	US-09-027-169-4	Sequence 4, Appli
4	3851.6	57.2	4476	US-08-801-344-2	Sequence 2, Appli
5	3851.6	57.2	4476	US-09-498-599-2	Sequence 2, Appli
6	2858	42.5	4557	US-08-778-717-5	Sequence 5, Appli
7	2729	40.6	4357	US-09-795-872-10	Sequence 10, Appli
8	2598.8	38.6	4593	US-08-801-344-1	Sequence 1, Appli
9	2598.8	38.6	4593	US-09-498-599-1	Sequence 1, Appli
10	2596	38.6	6312	US-08-531-601-3	Sequence 3, Appli
11	2596	38.6	6312	US-08-859-032-3	Sequence 3, Appli
12	2591.8	38.5	5238	5453363-1	Patent No. 5453363
13	2591.8	38.5	5238	5453363-1	Patent No. 5453363
14	2483.6	36.9	4410	US-08-594-469-1	Sequence 1, Appli
15	2483.6	36.9	4410	US-08-906-957-1	Sequence 1, Appli
16	2438.8	36.3	5446	US-09-358-856C-1	Sequence 1, Appli
17	2198.4	32.5	7892	US-07-916-098A-40	Sequence 40, Appli
18	2185.6	32.5	7633	US-09-028-851-1	Sequence 1, Appli
19	2185.6	32.5	7633	US-08-815-520-1	Sequence 1, Appli
20	2185.6	32.5	7633	US-09-273-163-1	Sequence 1, Appli
21	2183.4	32.4	8068	US-08-809-513A-7	Sequence 7, Appli
22	2183.4	32.4	8068	US-09-301-593-27	Sequence 27, Appli
23	2183.4	32.4	8068	US-09-301-593-35	Sequence 35, Appli
24	2183.4	32.4	9144	US-08-556-978B-79	Sequence 79, Appli
25	2182.6	32.4	7731	US-09-301-593-29	Sequence 29, Appli
26	2182.6	32.4	7731	US-09-301-593-42	Sequence 42, Appli
27	2182	32.4	3699	US-08-053-131-120	Sequence 120, App

ALIGNMENTS

RESULT 1

US-09-183-188B-5
; Sequence 5, Application US/09183188B
; Patent No. 6491905
; GENERAL INFORMATION:
; APPLICANT: Sorscher, Eric-J.
; APPLICANT: Parker, William B.
; APPLICANT: Waud, William
; APPLICANT: Gadi, Vijayakrishna K.
; TITLE OF INVENTION: RECOMBINANT E. COLI FOR DELIVERY OF PNP TO TUMOR CELLS
; FILE REFERENCE: UAB-12405/22
; CURRENT APPLICATION NUMBER: US/09/183,188B
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 08/702,181
; PRIOR FILING DATE: 1996-08-23
; PRIOR APPLICATION NUMBER: 08/122,321
; PRIOR FILING DATE: 1993-09-14
; PRIOR APPLICATION NUMBER: 60/064,676
; PRIOR FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 5013
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-183-188B-5

Query Match	57.3%	Score 3853.4;	DB 4;	Length 5013;
Best Local Similarity	100.0%;	Pred. No. 0;	Mismatches 1;	Indels 0;
Matches 3854;	Conservative	0;		Gaps 0;
QY	2875	AGCTTGGCTGTTTTCGCGGATGAGAGAAGATTTTCAGCTGTATACAGATTAAATCAGAAC	2934	
Db	1159	AGCTTGGCTGTTTTCGCGGATGAGAGAAGATTTTCAGCTGTATACAGATTAAATCAGAAC	1218	
QY	2935	GCAGAAGCGGTGTGATAAAACAGAAATTCCTCGCGGAGTAGCGGGTGTCCACCTG	2994	
Db	1219	GCAGAAGCGGTGTGATAAAACAGAAATTCCTCGCGGAGTAGCGGGTGTCCACCTG	1278	
QY	2995	ACCCCATCCGAACCTCAGAACTGAAACCGCTAGCGCGGATGAGTGTGGGGTCTCCCC	3054	
Db	1279	ACCCCATCCGAACCTCAGAACTGAAACCGCTAGCGCGGATGAGTGTGGGGTCTCCCC	1338	
QY	3055	ATCGCAGAGTAGGGAATCCAGGCGATCAAAATATAAAGAGGCTCAGTCGAAAGACTGG	3114	
Db	1339	ATCGCAGAGTAGGGAATCCAGGCGATCAAAATATAAAGAGGCTCAGTCGAAAGACTGG	1398	
QY	3115	GCCTTTCGTTTATCTGTTGTTGTCGTAACCGTCTCCTGAGTAGGAGCAAAATCGCGG	3174	
Db	1399	GCCTTTCGTTTATCTGTTGTTGTCGTAACCGTCTCCTGAGTAGGAGCAAAATCGCGG	1458	

QY 3175 GGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGAGGGTGGCGGCGAGGACGCGCCGCA 3234
 Db 1459 GGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGAGGGTGGCGGCGAGGACGCGCCGCA 1518
 QY 3235 TAAACTGCGCAGCATCAAAATTAAGCAGAGGCACTCTGACGGATGGCCCTTTTTCGGTTT 3294
 Db 1519 TAAACTGCGCAGCATCAAAATTAAGCAGAGGCACTCTGACGGATGGCCCTTTTTCGGTTT 1578
 QY 3295 CTACAAACTCTTTTGTGTTTATTTTCTAAATACATTTCAAAATATGATCCGCTCATGAGAC 3354
 Db 1579 CTACAAACTCTTTTGTGTTTATTTTCTAAATACATTTCAAAATATGATCCGCTCATGAGAC 1638
 QY 3355 AATAACCTGATAAATGCTTCAATTAATTAAGAAAGAGATGATGATTTCAACATT 3414
 Db 1639 AATAACCTGATAAATGCTTCAATTAATTAAGAAAGAGATGATGATTTCAACATT 1698
 QY 3415 TCCGTGTCGCCCTTATTCCTTTTGTGGGCAATTTGCTTCCCTGTTTTCCTCACCCAG 3474
 Db 1699 TCCGTGTCGCCCTTATTCCTTTTGTGGGCAATTTGCTTCCCTGTTTTCCTCACCCAG 1758
 QY 3475 AAAACCTGCTGAAAGTAAAGATGCTGAAGATCAGTTGGGTGCACGATGGGTACATCG 3534
 Db 1759 AAAACCTGCTGAAAGTAAAGATGCTGAAGATCAGTTGGGTGCACGATGGGTACATCG 1818
 QY 3535 AACTGGATCTCAACAGCGGTAAAGTCTTTGAGAGATTTTTCGCCCGCAAGACGTTTTCCAA 3594
 Db 1819 AACTGGATCTCAACAGCGGTAAAGTCTTTGAGAGATTTTTCGCCCGCAAGACGTTTTCCAA 1878
 QY 3595 TGATGAGCATTTTAAAGTCTGCTATGTCGGCGGTATATTCCTGTTGAGCTGACCGCGGC 3654
 Db 1879 TGATGAGCATTTTAAAGTCTGCTATGTCGGCGGTATATTCCTGTTGAGCTGACCGCGGC 1938
 QY 3655 AAGACAACTCGGTGCGCGCATACACTATTTCTCAGAATGACTTGGTTGAGTACTCACCG 3714
 Db 1939 AAGACAACTCGGTGCGCGCATACACTATTTCTCAGAATGACTTGGTTGAGTACTCACCG 1998
 QY 3715 TCAAGAAAAGCATCTTAAGGATGCGATGACAGTAAGAGAAATTAATGACGATGCTGCCATAA 3774
 Db 1999 TCACAGAAAAGCATCTTAAGGATGCGATGACAGTAAGAGAAATTAATGACGATGCTGCCATAA 2058
 QY 3775 CCATGAGTGAATAACCTCGGCCAACTTTACTTCTGACAAACGATCGGAGGAGC 3834
 Db 2059 CCATGAGTGAATAACCTCGGCCAACTTTACTTCTGACAAACGATCGGAGGAGC 2118
 QY 3835 TAAACGCTTTTGTGCAAACTGCGGATCATGTAACCTCGCCTTGATCGTTGGGAACCGG 3894
 Db 2119 TAAACGCTTTTGTGCAAACTGCGGATCATGTAACCTCGCCTTGATCGTTGGGAACCGG 2178
 QY 3895 AGCTGAATGAAGCCATACAAACGACGAGCGTGACACCAAGATGCTGTAGCAATGGCAA 3954
 Db 2179 AGCTGAATGAAGCCATACAAACGACGAGCGTGACACCAAGATGCTGTAGCAATGGCAA 2238
 QY 3955 CAACGTTGCGCAAACTTAACTGCGCACTTACTTACTAGCTTCCCGGCAACATTA 4014
 Db 2239 CAACGTTGCGCAAACTTAACTGCGCACTTACTTACTAGCTTCCCGGCAACATTA 2298
 QY 4015 TAGACTGATGAGCGGATAAAGTTGACGACCACTTTCTGCGCTCGGCCCTTTCGGCTG 4074
 Db 2299 TAGACTGATGAGCGGATAAAGTTGACGACCACTTTCTGCGCTCGGCCCTTTCGGCTG 2358
 QY 4075 GCTGTTTATTTGCTGATAAATCTGAGCGGTGAGCGTGGGTCTCGCGGTATCATTTGAG 4134
 Db 2359 GCTGTTTATTTGCTGATAAATCTGAGCGGTGAGCGTGGGTCTCGCGGTATCATTTGAG 2418
 QY 4135 CACTGGGCGCATGGTAAAGCCCTCCCGTATCGTAGTTATCTACACGAGCGGAGTCAGG 4194
 Db 2419 CACTGGGCGCATGGTAAAGCCCTCCCGTATCGTAGTTATCTACACGAGCGGAGTCAGG 2478
 QY 4195 CAACTATGGATGAACGAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCAT 4254
 Db 2479 CAACTATGGATGAACGAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCAT 2538

QY 4255 GGTAACTGTGACGACGAGTTTACTCATATATACATTTAGATTGATTTAAAACTTCATTTT 4314
 Db 2539 GGTAACTGTGACGACGAGTTTACTCATATATACATTTAGATTGATTTAAAACTTCATTTT 2598
 QY 4315 AATTTAAAGGATCTAGGTGAAGATCCTTTTGTGATATCTCATGACCAAAATCCCTTAAC 4374
 Db 2599 AATTTAAAGGATCTAGGTGAAGATCCTTTTGTGATATCTCATGACCAAAATCCCTTAAC 2658
 QY 4375 GTGAGTTTTCCTTCCACTGAGCGTCAGACCCGCTAGAAAAGATCAAAAGATCTTCTTGAG 4434
 Db 2659 GTGAGTTTTCCTTCCACTGAGCGTCAGACCCGCTAGAAAAGATCAAAAGATCTTCTTGAG 2718
 QY 4435 ATCCTTTTTCCTGCGGTATCTGCTTGTGCAAAACAAAACCCAGCTACCGCGG 4494
 Db 2719 ATCCTTTTTCCTGCGGTATCTGCTTGTGCAAAACAAAACCCAGCTACCGCGG 2778
 QY 4495 TGGTTTGTTCCTCCGATCAAGAGCTACCAACTCTTTTCCGAAGTAACTGGCTTCAGCA 4554
 Db 2779 TGGTTTGTTCCTCCGATCAAGAGCTACCAACTCTTTTCCGAAGTAACTGGCTTCAGCA 2838
 QY 4555 GAGCGAGATACCAAAATCTGCTCTCTAGTGTAGCGGTAGTTAGGCCACCACTTCAAGA 4614
 Db 2839 GAGCGAGATACCAAAATCTGCTCTCTAGTGTAGCGGTAGTTAGGCCACCACTTCAAGA 2898
 QY 4615 ACTCTGTAGCACCGCTACATACCTCGCTCTGCTAATCTCTTACCAAGTGGCTGCGCA 4674
 Db 2899 ACTCTGTAGCACCGCTACATACCTCGCTCTGCTAATCTCTTACCAAGTGGCTGCGCA 2958
 QY 4675 GTGCGATTAAGTGTCTTACCGGTTTGAAGTCAAGACGATAGTTACCGGATAGGCGC 4734
 Db 2959 GTGCGATTAAGTGTCTTACCGGTTTGAAGTCAAGACGATAGTTACCGGATAGGCGC 3018
 QY 4735 AGCGGTGCGGTGAACCGGGGTTCTGTGCAACAGCCAGCTTGGAGCGAACGACTACA 4794
 Db 3019 AGCGGTGCGGTGAACCGGGGTTCTGTGCAACAGCCAGCTTGGAGCGAACGACTACA 3078
 QY 4795 CCGAAGTGAATACCTACAGCGTGAGCTACAGAAAGCGCCAGCTTCCGAAAGGAGAA 4854
 Db 3079 CCGAAGTGAATACCTACAGCGTGAGCTACAGAAAGCGCCAGCTTCCGAAAGGAGAA 3138
 QY 4855 AGGCGGACAGGATATCCGGTAAGCGCGGTCGGAAACAGGAGAGCGCACGAGGAGCTTC 4914
 Db 3139 AGGCGGACAGGATATCCGGTAAGCGCGGTCGGAAACAGGAGAGCGCACGAGGAGCTTC 3198
 QY 4915 CAGGGGAAACGCTGTGATCTTTATAGTCTGTGGGTTTCGCGACCTCTGATTTGAGC 4974
 Db 3199 CAGGGGAAACGCTGTGATCTTTATAGTCTGTGGGTTTCGCGACCTCTGATTTGAGC 3258
 QY 4975 GTGATTTTGTGATGCTCGTCAGGGGGCGGAGCCTATGGAACAAACCGCAGCAACGCGG 5034
 Db 3259 GTGATTTTGTGATGCTCGTCAGGGGGCGGAGCCTATGGAACAAACCGCAGCAACGCGG 3318
 QY 5035 CTTTTCACGGTCTCTGGCTTTTGTGCTTGTCTCATATGTTCTTTCTCGCTTAT 5094
 Db 3319 CTTTTCACGGTCTCTGGCTTTTGTGCTTGTCTCATATGTTCTTTCTCGCTTAT 3378
 QY 5095 CCCCTGATTTCTGTGATAACCGTATTAACCGCTTTGAGTGAGTGAACCGCTCGCGCA 5154
 Db 3379 CCCCTGATTTCTGTGATAACCGTATTAACCGCTTTGAGTGAGTGAACCGCTCGCGCA 3438
 QY 5155 GCCGAAACGACGAGCGCAGGTCAGTGACGAGGAGGAGGAGCGCCTGATGCGGT 5214
 Db 3439 GCCGAAACGACGAGCGCAGGTCAGTGACGAGGAGGAGGAGGAGCGCCTGATGCGGT 3498
 QY 5215 ATTTTCTCTTACGCATCTGTGCGGTATTTTCAACCGCATATGTTGCACTCTCAGTACAA 5274
 Db 3499 ATTTTCTCTTACGCATCTGTGCGGTATTTTCAACCGCATATGTTGCACTCTCAGTACAA 3558
 QY 5275 TCTGCTGTGATGCGCGCATAGTTAAGCCAGTATACCTCCGCTATCGCTACGTCAGTGGT 5334
 Db 3559 TCTGCTGTGATGCGCGCATAGTTAAGCCAGTATACCTCCGCTATCGCTACGTCAGTGGT 3618
 QY 5335 CATGCTGCGCCCGACACCCGCCCAACCCGCTGACGCGCCTGACGCGGCTTGTCTGCT 5394

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 4, 2005, 15:42:19 ; Search time 3779 Seconds
(without alignments)

11542.622 Million cell updates/sec

Title: US-10-649-433C-1

Perfect score: 6729

Sequence: 1 gtttagacgcttatcatcga.....agtttagcggaattgatctg 6729

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 7297361 seqs, 3241162794 residues

Total number of hits satisfying chosen parameters: 14594722

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US05_PUBCOMB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US04_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US03_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/US02_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US01_PUBCOMB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US00_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09D_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
- 19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
- 20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
- 21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
- 22: /cgn2_6/ptodata/1/pubpna/US10J_PUBCOMB.seq:*
- 23: /cgn2_6/ptodata/1/pubpna/US10K_PUBCOMB.seq:*
- 24: /cgn2_6/ptodata/1/pubpna/US10L_PUBCOMB.seq:*
- 25: /cgn2_6/ptodata/1/pubpna/US10M_PUBCOMB.seq:*
- 26: /cgn2_6/ptodata/1/pubpna/US10N_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6729	100.0	6729	21	US-10-649-433C-1
2	3853.4	57.3	5013	14	US-10-198-034-5
3	3853.4	57.3	5013	15	US-10-198-228-5
4	3852.2	57.2	5926	13	US-10-024-809-3
5	3852.2	57.2	5926	13	US-10-024-809-4
6	3850.6	57.2	4176	16	US-10-244-142A-3
7	3344.8	49.7	5024	17	US-10-313-963A-52

C	8	3344.8	49.7	5448	21	US-10-671-995A-2	Sequence 2, Appli
C	9	3344.8	49.7	5454	21	US-10-671-995A-1	Sequence 1, Appli
	10	3341.2	49.7	6464	15	US-10-151-690-20	Sequence 20, Appli
	11	2230.2	48.0	5398	22	US-10-505-082-3	Sequence 3, Appli
	12	2877.4	42.8	5522	10	US-09-872-868-16	Sequence 16, Appli
	13	2877.4	42.8	5525	10	US-09-872-868-17	Sequence 17, Appli
	14	2877.4	42.8	6086	10	US-09-872-868-12	Sequence 12, Appli
	15	2877.4	42.8	6086	10	US-09-872-868-13	Sequence 13, Appli
	16	2877.4	42.8	6086	10	US-09-872-868-14	Sequence 14, Appli
	17	2877.4	42.8	6088	10	US-09-872-868-15	Sequence 15, Appli
	18	2863.2	42.6	5451	20	US-10-485-281-6	Sequence 6, Appli
	19	2858	42.5	4557	17	US-10-457-372-5	GENERAL INFORMA
	20	2858	42.5	17135	10	US-09-843-150-55	Sequence 55, Appli
	21	2858	42.5	17135	20	US-10-475-962-55	Sequence 55, Appli
	22	2757.2	41.0	4700	17	US-10-313-963A-54	Sequence 54, Appli
	23	2743.4	40.8	5772	20	US-10-131-998A-37	Sequence 37, Appli
	24	2743.4	40.8	5772	20	US-10-131-998A-38	Sequence 38, Appli
	25	2731	40.6	4145	20	US-10-699-088-163	Sequence 163, App
	26	2731	40.6	4145	21	US-10-699-114-163	Sequence 163, App
	27	2729	40.6	4357	10	US-09-795-872-10	Sequence 10, Appli
	28	2729	40.6	4357	19	US-10-662-824-10	Sequence 20, Appli
	29	2664.2	39.6	6886	18	US-10-466-642-20	Sequence 20, Appli
	30	2664.2	39.6	6886	18	US-10-466-717-20	Sequence 20, Appli
	31	2658.4	39.5	5797	17	US-10-316-194-167	Sequence 167, App
	32	2658.4	39.5	5797	21	US-10-732-180-203	Sequence 203, App
	33	2613.4	38.8	6806	10	US-09-908-943A-194	Sequence 194, App
	34	2613.4	38.8	6806	20	US-10-801-487-194	Sequence 194, App
	35	2613.4	38.8	6806	20	US-10-801-938-194	Sequence 194, App
	36	2613.4	38.8	6806	20	US-10-801-509-194	Sequence 194, App
	37	2613.4	38.8	6806	20	US-10-801-486-194	Sequence 194, App
	38	2613.4	38.8	6806	21	US-10-801-493-194	Sequence 194, App
	39	2603.2	38.7	4921	21	US-10-495-491-2	Sequence 2, Appli
	40	2603.2	38.7	4945	21	US-10-495-491-3	Sequence 3, Appli
	41	2600.4	38.6	4935	21	US-10-495-491-1	Sequence 1, Appli
	42	2600.4	38.6	4951	21	US-10-495-491-4	Sequence 4, Appli
	43	2464.2	36.6	4956	16	US-10-227-617A-5	Sequence 5, Appli
	44	2439.8	36.3	5446	17	US-10-393-269-1	Sequence 1, Appli
C	45	2369	35.2	7615	10	US-09-898-616A-9	Sequence 9, Appli

ALIGNMENTS

RESULT 1
US-10-649-433C-1
; Sequence 1, Application US/10649433C
; Publication No. US20050118726A1
; GENERAL INFORMATION:
; APPLICANT: Schultz, Jerome S.
; TITLE OF INVENTION: System and Method for Detecting Bioanalytes and Method for
; TITLE OF INVENTION: Producing a Bioanalyte Sensor
; FILE REFERENCE: 03-016
; CURRENT APPLICATION NUMBER: US/10/649,433C
; CURRENT FILING DATE: 2003-08-26
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 6729
; TYPE: DNA
; ORGANISM: Escherichia coli
; US-10-649-433C-1

Query Match 100.0%; Score 6729; DB 21; Length 6729;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	GTTTGACAGCTTATCATCGCTGACCGTGACCAATGCTTCTGGCGTCAGGCAGCCATC	60
Db	1	GTTTGACAGCTTATCATCGCTGACCGTGACCAATGCTTCTGGCGTCAGGCAGCCATC	60
Qy	61	CGAAGCTGTGTATGGCTGTGTCAGGTCTTAATCACTCATTAATTCGTCTCGCTCAAGC	120
Db	61	CGAAGCTGTGTATGGCTGTGTCAGGTCTTAATCACTCATTAATTCGTCTCGCTCAAGC	120

QY 121 GCACTCCGTTCTGATTAATGTTTTTGGCCGACATCATAAAGGTTCTGCAAAATATTC 180
DB 121 GCACTCCGTTCTGATTAATGTTTTTGGCCGACATCATAAAGGTTCTGCAAAATATTC 180
QY 181 TGAATGAGCTGTTGACAAATTAATCATCCGCTCGTATTAATGTTGGAAATTTGTGAGCGGA 240
DB 181 TGAATGAGCTGTTGACAAATTAATCATCCGCTCGTATTAATGTTGGAAATTTGTGAGCGGA 240
QY 241 TAAACAATTTACACAGGAAACAGCCCGCTGAGAAAAAGCAAGCGGCACCTGCTCTTTAA 300
DB 241 TAAACAATTTACACAGGAAACAGCCCGCTGAGAAAAAGCAAGCGGCACCTGCTCTTTAA 300
QY 301 CAAATTTATCAGCAATCTGTGTGGCACTCGACCGGAATTAATCGAATTAATTTATTA 360
DB 301 CAAATTTATCAGCAATCTGTGTGGCACTCGACCGGAATTAATCGAATTAATTTATTA 360
QY 361 AAAATTAAGAGGTATATTAATGATATCGAATTAATTAAGGAGGAATAAACATGTTGAG 420
DB 361 AAAATTAAGAGGTATATTAATGATATCGAATTAATTAAGGAGGAATAAACATGTTGAG 420
QY 421 CAAGGGCAGGAGCTGTTCAACCGGGTGGTCCCATCTCGTTCGAGCTGGACGGCGAGCT 480
DB 421 CAAGGGCAGGAGCTGTTCAACCGGGTGGTCCCATCTCGTTCGAGCTGGACGGCGAGCT 480
QY 481 AAAGGGCACAAGTTTCAGCTGTCGGCGAGGGCGAGGGCGATGCGACCTACGGCAAGCT 540
DB 481 AAAGGGCACAAGTTTCAGCTGTCGGCGAGGGCGAGGGCGATGCGACCTACGGCAAGCT 540
QY 541 GACCTGGAAGTTCACTGCAACCGGCAAGCTGCCGCTGCTGCGGCGGCGGCGGCGGCGTAC 600
DB 541 GACCTGGAAGTTCACTGCAACCGGCAAGCTGCCGCTGCTGCGGCGGCGGCGGCGGCGTAC 600
QY 601 CACCTTCGGCTACGGCTGCAAGTCTTCGGCGGCTTACCCCGGCTACCCCGACACATGAAGCAGCAGA 660
DB 601 CACCTTCGGCTACGGCTGCAAGTCTTCGGCGGCTTACCCCGGCTACCCCGACACATGAAGCAGCAGA 660
QY 661 CTTCTTCAAGTCCGCGCATGCGGCAAGGCTAGTTCAGGAGGCGGCGGCGGCGGCGGCGGCGG 720
DB 661 CTTCTTCAAGTCCGCGCATGCGGCAAGGCTAGTTCAGGAGGCGGCGGCGGCGGCGGCGGCGG 720
QY 721 CGAGGCAACTACAGACCCGCGCGAGTTCGAGTTCGAGGCGGCGGCGGCGGCGGCGGCGGCGG 780
DB 721 CGAGGCAACTACAGACCCGCGCGAGTTCGAGTTCGAGGCGGCGGCGGCGGCGGCGGCGGCGG 780
QY 781 CATCGAGCTGAAGGCGCATCGACTTCAAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
DB 781 CATCGAGCTGAAGGCGCATCGACTTCAAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
QY 841 GTACAACTACAGCAGCAGAGCTCTATCATGCGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGG 900
DB 841 GTACAACTACAGCAGCAGAGCTCTATCATGCGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGG 900
QY 901 GGTGAACCTTCAAGATCCGCGCAACATCGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
DB 901 GGTGAACCTTCAAGATCCGCGCAACATCGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
QY 961 CCAGCAGAACACCCCATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
DB 961 CCAGCAGAACACCCCATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
QY 1021 CTACAGTCCGCGCTGAGCAAGAGCCCAACGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
DB 1021 CTACAGTCCGCGCTGAGCAAGAGCCCAACGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
QY 1081 GTTCGTGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
DB 1081 GTTCGTGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
QY 1141 TACTCGCAATGGTGTAAACAATCTAATAGTACGAGCAATCTTAATGCTGTAGTGGCGCAA 1200
DB 1141 TACTCGCAATGGTGTAAACAATCTAATAGTACGAGCAATCTTAATGCTGTAGTGGCGCAA 1200

QY 1201 GGCTATTGACGACGATGCGAAAGCGCGCCAGATGTTTCAGCTGCTGATGAATGATTCTCA 1260
DB 1201 GGCTATTGACGACGATGCGAAAGCGCGCCAGATGTTTCAGCTGCTGATGAATGATTCTCA 1260
QY 1261 GAATGACAGTCCAGGACGACGATCAGATCGAGTATTGCTGGCCAAAGGGGTGAAGGC 1320
DB 1261 GAATGACAGTCCAGGACGACGATCAGATCGAGTATTGCTGGCCAAAGGGGTGAAGGC 1320
QY 1321 ACTGCCCATCAACTGGTTGACCCGGCAGCTGCGGGTACCGGTGATTGAGAAAGCGCGTGG 1380
DB 1321 ACTGCCCATCAACTGGTTGACCCGGCAGCTGCGGGTACCGGTGATTGAGAAAGCGCGTGG 1380
QY 1381 GCAAAAAGTCCGCGGTGTTTCTTCAACAAAGAACCGTCTCTGTAAGCGCTGATAGCTA 1440
DB 1381 GCAAAAAGTCCGCGGTGTTTCTTCAACAAAGAACCGTCTCTGTAAGCGCTGATAGCTA 1440
QY 1441 CGACAAAGCCTTACTGTTGGCACTGACTCAAAAGAGTCCGGCATTTATTCAAGGGCGATT 1500
DB 1441 CGACAAAGCCTTACTGTTGGCACTGACTCAAAAGAGTCCGGCATTTATTCAAGGGCGATT 1500
QY 1501 GATTGCTAAAACA CTGGGCGGCGAATCAGGGTTGGGATCTGAACAAAGACGGTCAAGATTCA 1560
DB 1501 GATTGCTAAAACA CTGGGCGGCGAATCAGGGTTGGGATCTGAACAAAGACGGTCAAGATTCA 1560
QY 1561 GTTCGTACTGCTGAAAGGTGAACCGGGCCATCCGGATGCGAGACGATCACTTAAGT 1620
DB 1561 GTTCGTACTGCTGAAAGGTGAACCGGGCCATCCGGATGCGAGACGATCACTTAAGT 1620
QY 1621 GATTAAAGAAATTGAACGATTAAGGATCAAAATCTGAACAGTTACAGTTAGATACCGCAAT 1680
DB 1621 GATTAAAGAAATTGAACGATTAAGGATCAAAATCTGAACAGTTACAGTTAGATACCGCAAT 1680
QY 1681 GTGGGACACCGCTCAGGCGAAAGATAAGATGGAACGCTGGCTGTCTGCCCGAACGCCAA 1740
DB 1681 GTGGGACACCGCTCAGGCGAAAGATAAGATGGAACGCTGGCTGTCTGCCCGAACGCCAA 1740
QY 1741 CAAAATCGAAGTGGTTATCGCCAAACGATGCGATGCGCATGGCGCGGTTGAGCGCT 1800
DB 1741 CAAAATCGAAGTGGTTATCGCCAAACGATGCGATGCGCATGGCGCGGTTGAGCGCT 1800
QY 1801 GAAAGCACACAAAGTCCAGCATTCGGTGTGTTGGGCTCGATGCGCTGCGCAGAACGCGCT 1860
DB 1801 GAAAGCACACAAAGTCCAGCATTCGGTGTGTTGGGCTCGATGCGCTGCGCAGAACGCGCT 1860
QY 1861 GGCCTGTGTAATCCGGTGCATCTGGGCGGCAACCGTACTGAAACGATGCTAAACACAGGC 1920
DB 1861 GGCCTGTGTAATCCGGTGCATCTGGGCGGCAACCGTACTGAAACGATGCTAAACACAGGC 1920
QY 1921 GAAAGCGACCTTTCATCTGGCGAAACCTGCGCGATGTTAAAGTGGCGCTGATGGCAC 1980
DB 1921 GAAAGCGACCTTTCATCTGGCGAAACCTGCGCGATGTTAAAGTGGCGCTGATGGCAC 1980
QY 1981 CAACTGGAATAATCGACAAACAAAGTGGTCCGCTACCTTATGTTGGCGTAGATAAGACAA 2040
DB 1981 CAACTGGAATAATCGACAAACAAAGTGGTCCGCTACCTTATGTTGGCGTAGATAAGACAA 2040
QY 2041 CCTGGCTGAATTCAGCAAGAAAGGTACAGTAAAGGAGAAACCTTTTCACTGGAGTTGT 2100
DB 2041 CCTGGCTGAATTCAGCAAGAAAGGTACAGTAAAGGAGAAACCTTTTCACTGGAGTTGT 2100
QY 2101 CCCAATTCCTGTTGAAATAGATGGTGTGTTAAATGGGCAACAAATTTCTGTCAGTGGAGA 2160
DB 2101 CCCAATTCCTGTTGAAATAGATGGTGTGTTAAATGGGCAACAAATTTCTGTCAGTGGAGA 2160
QY 2161 GGGTGAAGGTGATGCAACATACGGAACCTTACCTTTAAATTTATTTGCACTACTGGAAA 2220
DB 2161 GGGTGAAGGTGATGCAACATACGGAACCTTACCTTTAAATTTATTTGCACTACTGGAAA 2220
QY 2221 ACTACCTGTTCCATGGCCAAACCTTTGTCATCTTTCTTTATGGTGTGTTCAATGCTTTTC 2280
DB 2221 ACTACCTGTTCCATGGCCAAACCTTTGTCATCTTTCTTTATGGTGTGTTCAATGCTTTTC 2280
QY 2281 CCGTTATCCGGATCATATGAAACGGCATGACTTTTTCAAGAGTGCCTGCCGAGGTTA 2340

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 2, 2005, 18:19:33 ; Search time 201.5 Seconds
(without alignments)
6426.229 Million cell updates/sec

Title: US-10-649-433C-1
Perfect score: 11920
Sequence: 1 gtttagacgttatcatcg.....agttagcgcaattgatctg 6729

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delopt 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/cgn2.1/USPTO spool_p/US10649433/runat_02082005_164945_1532/app_query.fasta_1.6919
-DB=PIR_QPWT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10649433 @CNC 1 1 332 @runat_02082005_164945_1532 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1765	14.8	360	1 RPCL	lactose operon rep
2	1765	14.8	360	4 S58116	lac repressor - sy
3	1754	14.7	360	2 F90678	lac operon transcr
4	1747	14.7	360	2 B85529	lac operon transcr
5	1579	13.2	332	1 JGSCG	D-galactose-binding
6	1579	13.2	332	2 B91009	galactose-binding
7	1579	13.2	332	2 D85853	galactose-binding
8	1537	12.9	332	1 S15554	D-galactose-binding
9	1510	12.7	332	2 AF0781	D-galactose-binding
10	1508	12.7	332	1 S29390	galactose-binding
11	1467	12.3	286	2 T51301	beta-lactamase (EC
12	1467	12.3	286	2 S47061	beta-lactamase (EC
13	1467	12.3	286	4 S41975	beta-lactamase (EC
14	1464	12.3	286	1 PNECP	beta-lactamase (EC

15	1464	12.3	286	4 I40905	beta-lactamase (EC
16	1451	12.2	286	2 S60312	extended spectrum
17	1450	12.2	286	2 S60313	beta-lactamase (EC
18	1444	12.1	286	2 S60310	extended spectrum
19	1442	12.1	286	2 S60311	beta-lactamase (EC
20	1438	12.1	286	2 JQ1546	Bla protein - Salm
21	1425.5	12.0	335	2 AG0183	galactose-binding
22	1242	10.4	238	1 JQ1514	green-fluorescent
23	1167.5	9.8	324	2 C82214	galactoside ABC tr
24	1137.5	9.5	349	1 G64096	D-galactose-binding
25	960	8.1	286	2 S16146	beta-lactamase (EC
26	960	8.1	286	2 A60679	beta-lactamase (EC
27	956	8.0	286	1 A44998	beta-lactamase (EC
28	953	8.0	286	2 A37200	beta-lactamase (EC
29	952	8.0	286	2 A60632	beta-lactamase (EC
30	937	7.9	265	2 S04464	beta-lactamase (EC
31	931	7.8	265	2 S02434	beta-lactamase (EC
32	927.5	7.8	287	1 A44996	beta-lactamase (EC
33	924	7.8	265	2 A60448	beta-lactamase (EC
34	917	7.7	279	2 A24469	beta-lactamase (EC
35	896	7.5	286	1 A44958	beta-lactamase (EC
36	715	6.0	357	2 AE0104	lactose operon rep
37	674	5.6	149	2 JQ1541	hypothetical 16.9K
38	669.5	5.6	298	2 A41381	beta-lactamase (EC
39	632	5.3	354	2 B24925	lac repressor - Kl
40	540.5	4.5	304	2 A35001	beta-lactamase (EC
41	536.5	4.5	304	2 A49789	beta-lactamase (EC
42	530.5	4.5	288	2 JQ1136	beta-lactamase (EC
43	524	4.4	105	2 JC2566	bla protein - pseu
44	519.5	4.4	281	2 D95395	probable Beta lact
45	518.5	4.3	288	2 JS0755	beta-lactamase (EC

ALIGNMENTS

RESULT 1

RPECL
lactose operon repressor - Escherichia coli (strain K-12)
N:Alternate names: lac repressor
C:Species: Escherichia coli
C:Date: 30-Apr-1982 #sequence revision 30-Apr-1982 #text change 09-Jul-2004
C:Accession: A93198; A91234; A92122; A93785; S40661; S02540; A64762; S68009; S14614; A03
R:Farabaugh, P.J.
Nature 274, 765-769, 1978
A:Title: Sequence of the lacI gene.
A:Reference number: A93198; MUID:78246991; PMID:355891
A:Accession: A93198
A:Molecule type: DNA
A:Residues: 1-360 <FAR>
A:Cross-references: UNIPROT:P03023
R:Beyreuther, K.; Adler, K.; Fanning, E.; Murray, A.; Geisler, N.
Eur. J. Biochem. 59, 491-509, 1975
A:Title: Amino-acid sequence of lac repressor from Escherichia coli. Isolation, sequence
A:Reference number: A91234; MUID:76091932; PMID:1107032
A:Accession: A91234
A:Molecule type: protein
A:Residues: 1-147;159-163,'O',165-230;233-360 <BEY>
A:Note: the active repressor is a tetramer of identical chains
A:Note: this protein was obtained from a strain with the I-SQ mutation, which leads to a
R:Beyreuther, K.
Nature 274, 767, 1978
A:Reference number: A93199
A:Contents: annotation; revision
A:Note: the revised sequence is identical with that shown
R:Platt, T.; Files, J.G.; Weber, K.
J. Biol. Chem. 248, 110-121, 1973
A:Reference number: A92122; MUID:73143730; PMID:4571224
A:Accession: A92122
A:Molecule type: protein
A:Residues: 1-59;96-101;206-215;328-347 <PLA>
A:Note: removal of residues 1-59 and 328-347 results in the molecule losing its DNA-bin
R:Ganem, D.; Miller, J.H.; Files, J.G.; Platt, T.; Weber, K.


```

RESULT 2
S58116
lac repressor - synthetic
C:Species: synthetic
C:Date: 13-Jan-1996 #sequence_revision 22-Aug-1996 #text_change 20-Oct-2000
C:Accession: S58116
R:Kahrs, A.F.; Odenbreit, S.; Schmitt, W.; Heuermann, D.; Meyer, T.F.; Haas, R.

```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 5, 2005, 01:59:21 ; Search time 19523 Seconds
(without alignments)
13119.627 Million cell updates/sec

Title: US-10-649-433C-1
Perfect score: 6729
Sequence: 1 gtttagcagcttatcatcga.....agttagcgaatgatctg 6729

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gse1:*
9: gb_gse2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	1055.8	15.7	1070	1	AJ281552
C 2	1018.2	15.1	1048	7	CO552396
C 3	984.8	14.6	1013	4	BM438846
C 4	924	13.7	1004	1	AJ281480
C 5	889.8	13.2	928	7	CO487414
C 6	845	12.6	902	7	CR753463
C 7	837.8	12.5	841	1	AL042026
C 8	835.8	12.4	935	4	BG838279
C 9	823	12.2	854	4	BM438950
C 10	820.2	12.2	865	7	CK125894
C 11	812.8	12.1	1126	8	BZ577702
C 12	806.4	12.0	819	6	CD649375
C 13	800.6	11.9	856	7	CN823189
C 14	796	11.8	1011	8	BZ576726
C 15	791.4	11.8	827	7	CN823392
C 16	785.4	11.7	1574	8	BZ572566
C 17	779.4	11.6	1073	7	CF269652
C 18	778	11.6	794	5	BQ751936
C 19	777.8	11.6	800	1	AJ281449
C 20	773.2	11.5	1336	8	BZ575810
C 21	772.8	11.5	789	6	CD280920
C 22	772	11.5	846	7	CV468077
C 23	770.6	11.5	779	7	CK394523
C 24	770.4	11.4	1249	8	BZ572284

25	769.2	11.4	832	7	CM822433
26	766.4	11.4	769	7	CK118014
27	764.2	11.4	769	7	CV224987
C 28	763.8	11.4	966	8	BZ570738
C 29	762.4	11.3	793	7	CK122681
C 30	762	11.3	793	5	BQ751655
31	761.6	11.3	780	5	BQ825693
32	758.2	11.3	820	7	CR753457
33	755.4	11.2	759	6	CD279661
34	755.4	11.2	806	5	BQ751220
35	753.8	11.2	1463	8	BZ571475
36	749.8	11.1	786	7	CN823164
37	747.4	11.1	752	7	CR766850
C 38	745.4	11.1	844	8	BZ574513
C 39	742	11.0	832	4	BG923768
C 40	741.4	11.0	758	9	CL422788
41	739.4	11.0	741	6	CD279174
C 42	739.2	11.0	914	8	BZ569398
43	733.2	10.9	811	9	ATH517156
44	733.2	10.9	1370	8	BZ571721
45	732.2	10.9	950	8	BZ571129

ALIGNMENTS

RESULT 1	AJ281552	1070 bp	mRNA	linear	EST 30-JUN-2000
LOCUS	4A3A-P6F11-F	Anopheles gambiae	immune competent	4A3A	Anopheles
DEFINITION	gambiae cDNA clone 4A3A-P6F11, mRNA sequence.				
ACCESSION	AJ281552				
VERSION	AJ281552.1	GI:6929432			
KEYWORDS	EST.				
SOURCE	Anopheles gambiae	(African malaria mosquito)			
ORGANISM	Anopheles gambiae				
REFERENCE	1 (bases 1 to 1070)				
AUTHORS	Dimopoulos, G., Casavant, T.L., Chang, S., Scheetz, T., Roberts, C., Donohue, M., Schultz, J., Benes, V., Bork, P., Ansong, W., Soares, M.B. and Kafatos, F.C.				
TITLE	Anopheles gambiae pilot gene discovery project: identification of mosquito innate immunity genes from expressed sequence tags generated from immune-competent cell lines				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	97 (12)			6619-6624 (2000)
MEDLINE	20300950				
PUBMED	10841561				
COMMENT	Contact: Dimopoulos G Fotis C. Kafatos laboratory European Molecular Biology Laboratory Meyerhofstrasse 1, 69117 Heidelberg, Germany. Location/Qualifiers 1. 1070 /organism="Anopheles gambiae" /mol_type="mRNA" /strain="4A r/r" /db_xref="taxon:7165" /clone="4A3A-P6F11" /cell_line="immune competent 4A3A" /lab_host="E. coli DH10B" /note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker; Site_1: EcoRI; Site_2: NotI; sequenced from forward priming site which reads from the 3' end of the cDNA. The 4A3A is a directionally cloned and normalized library that was constructed from the 4A3A cell line oligo-T primed cDNA according to: Bonaldo, Lennon & Soares (1996) : Normalization and Subtraction: Two approaches To Facilitate Gene Discovery, Genome Research 6, 791-806."				
FEATURES	source				
ORIGIN					

Query Match		15.7%;	Score 1055.8;	DB 1;	Length 1070;
Best Local Similarity		99.7%;	Pred. No. 1.2e-272;		
Matches 1068;		Conservative 0;	Mismatches 2;	Indels 1;	Gaps 1;
Qy	3959	GTTCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGCAACAACTAATAGA	4018		
Db	1070	GTTCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGCAACAACTAATAGA	1011		
Qy	4019	CTGGATGAGGCGGATAAAGTTGACAGCACTCTTCGCGCTGGCCCTTCGCGGTGGCTG	4078		
Db	1010	CTGGATGAGGCGGATAAAGTTGACAGCACTCTTCGCGCTGGC-CTTTCGCGGTGGCTG	952		
Qy	4079	GTTTATTGCTGATAAATCTGAGCCGCTGAGCGGTGGTCTGCGGTATCATTCGACACT	4138		
Db	951	GTTTATTGCTGATAAATCTGAGCCGCTGAGCGGTGGTCTGCGGTATCATTCGACACT	892		
Qy	4139	GGGGCCAGATGATAGCCCTCCGCTATCGTAGTTATCTACAGACGGGGAGTCAGGCAAC	4198		
Db	891	GGGGCCAGATGATAGCCCTCCGCTATCGTAGTTATCTACAGACGGGGAGTCAGGCAAC	832		
Qy	4199	TATGATGAACGAATAGACAGATCGCTGAGATAGTGCCTCACTGATTAAGCAATTGTA	4258		
Db	831	TATGATGAACGAATAGACAGATCGCTGAGATAGTGCCTCACTGATTAAGCAATTGTA	772		
Qy	4259	ACTGTCAGACCAAGTTTACTCATATATATCTTATGATTGATTTAAACTCTTATTAAT	4318		
Db	771	ACTGTCAGACCAAGTTTACTCATATATATCTTATGATTGATTTAAACTCTTATTAAT	712		
Qy	4319	TAAAGGATCTAGGTGAAGATCCTTTTGTATATCTCATGACCAAAATCCCTTAACGTGA	4378		
Db	711	TAAAGGATCTAGGTGAAGATCCTTTTGTATATCTCATGACCAAAATCCCTTAACGTGA	652		
Qy	4379	GTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAAGATCTTCTTGAGATCC	4438		
Db	651	GTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAAGATCTTCTTGAGATCC	592		
Qy	4439	TTTTTTCTCGCGTAATCTGCTGCTGCTGCACCAAAAACACCGCTACAGCGGTGGT	4498		
Db	591	TTTTTTCTCGCGTAATCTGCTGCTGCTGCACCAAAAACACCGCTACAGCGGTGGT	532		
Qy	4499	TGTTTTCGCGATCAAGAGCTACCAACTCTTTTCCGAGGTAACTGGCTTCAGCAGAGC	4558		
Db	531	TGTTTTCGCGATCAAGAGCTACCAACTCTTTTCCGAGGTAACTGGCTTCAGCAGAGC	472		
Qy	4559	CGAGATACAAATACTGCTCTTATGATGAGCGGTAGTTCAGCCACCACTTCAAGAACTC	4618		
Db	471	CGAGATACAAATACTGCTCTTATGATGAGCGGTAGTTCAGCCACCACTTCAAGAACTC	412		
Qy	4619	TGTAGCAGCGCTACATACCTCGCTCTGTAATCTGTTTACAGTGGCTGTCGCAAGTG	4678		
Db	411	TGTAGCAGCGCTACATACCTCGCTCTGTAATCTGTTTACAGTGGCTGTCGCAAGTG	352		
Qy	4679	CGATAAGTCGTGCTTTACCGGGTTCGAACTCAAGACGATAGTTTACCGGATAAGCGCAGCG	4738		
Db	351	CGATAAGTCGTGCTTTACCGGGTTCGAACTCAAGACGATAGTTTACCGGATAAGCGCAGCG	292		
Qy	4739	GTGCGGCTGAACGGGGGTTCGTGCACACAGCCAGCTTGGAGCGAAAGCACTACACCGA	4798		
Db	291	GTGCGGCTGAACGGGGGTTCGTGCACACAGCCAGCTTGGAGCGAAAGCACTACACCGA	232		
Qy	4799	ACTGAGATACCTACAGCGTGAGCTATGAGAAGCGCCAGCTTCCGGAAGGGAGAAAGGC	4858		
Db	231	ACTGAGATACCTACAGCGTGAGCTATGAGAAGCGCCAGCTTCCGGAAGGGAGAAAGGC	172		
Qy	4859	GGACAGGTATCCGGTAAGCGGAGGGTCCGAAACAGGAGAGCGCAGAGGAGCTTCCAGG	4918		
Db	171	GGACAGGTATCCGGTAAGCGGAGGGTCCGAAACAGGAGAGCGCAGAGGAGCTTCCAGG	112		
Qy	4919	GGGAAACCGCTGATCTTTATGATCTGTCGGGTTTCGCCACTCTGACTTGAGCGTCG	4978		
Db	111	GGGAAACCGCTGATCTTTATGATCTGTCGGGTTTCGCCACTCTGACTTGAGCGTCG	52		
Qy	4979	ATTTTGTGTGCTCGTCAGGGGGCGGAGCCCTATGGAAAAACGCCACAA	5029		

Db	51	ATTTTGTGATGCTCGTCAGGGGGCGGAGCCCTATGGAAAAACGCCAGCAA	1																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																						
----	----	--	---	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 2, 2005, 18:11:07 ; Search time 1027.5 Seconds
(without alignments)
6707.116 Million cell updates/sec

Title: US-10-649-433c-1
Perfect score: 11920
Sequence: 1 gtttgacagcttatcatcg.....agttagcgcgaattgatctg 6729

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0
Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 3224756

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-O=/cn2 1/USPTO spool P/US10649433/runat 02082005 164945 1521/app query.fasta_1.6919
-DB=UniProt -QFWT=fascan -SUFFIX=rup -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10649433 @CN 1 1 1271 @runat 02082005 164945 1521 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt 03.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1765	14.8	360	1 LACI_ECOLI	P03023 escherichia
2	1754	14.7	360	2 Q7AH57	Q7AH57 escherichia
3	1747	14.7	360	2 Q8X684	Q8X684 escherichia
4	1579	13.2	332	1 DGAL_ECOLI	P02927 escherichia
5	1579	13.2	332	2 Q8X641	Q8X641 escherichia
6	1537	12.9	332	1 DGAL_CITFR	P23925 citrobacter
7	1518	12.7	332	1 DGAL_SALTY	P23905 salmonella
8	1510	12.7	332	2 Q8Z5A4	Q8Z5A4 salmonella
9	1467	12.3	286	2 Q38058	Q38058 bacterioph
10	1467	12.3	286	2 Q00626	Q00626 staphylococ
11	1467	12.3	286	2 Q79CL6	Q79CL6 methylobaci
12	1467	12.3	286	2 Q79DR3	Q79DR3 escherichia
13	1464	12.3	286	1 BLAT_ECOLI	P62593 escherichia
14	1464	12.3	286	1 BLAT_SALTY	P62594 salmonella
15	1464	12.3	286	2 Q6A253	Q6A253 haemophilus
16	1464	12.3	286	2 Q61BN9	Q61BN9 pseudomonas

17	1464	12.3	286	2 Q6LCV6	Q6LCV6 neisseria g
18	1464	12.3	286	2 Q6TMH1	Q6TMH1 streptococc
19	1464	12.3	286	2 Q6W9J1	Q6W9J1 enterobacte
20	1464	12.3	286	2 Q6WRX2	Q6WRX2 zymomonas m
21	1464	12.3	286	2 Q6WZD4	Q6WZD4 acinetobact
22	1464	12.3	286	2 Q7B3X5	Q7B3X5 citrobacter
23	1464	12.3	286	2 Q7B899	Q7B899 klebsiella
24	1464	12.3	286	2 Q7BP57	Q7BP57 shigella fl
25	1464	12.3	286	2 Q7BR75	Q7BR75 neisseria m
26	1464	12.3	286	2 Q7DFY3	Q7DFY3 salmonella
27	1464	12.3	286	2 Q7DHD3	Q7DHD3 serratia ma
28	1464	12.3	286	2 Q799Y1	Q799Y1 plasmid ppv
29	1464	12.3	286	2 Q7BVP8	Q7BVP8 plasmid pet
30	1461	12.3	286	2 Q8KSD2	Q8KSD2 escherichia
31	1460	12.2	286	2 Q32372	Q32372 capnocytoph
32	1460	12.2	286	2 Q8KSD3	Q8KSD3 klebsiella
33	1460	12.2	286	2 Q932Y6	Q932Y6 pseudomonas
34	1460	12.2	286	2 Q93A80	Q93A80 escherichia
35	1460	12.2	286	2 Q6QQR1	Q6QQR1 proteus mir
36	1460	12.2	286	2 Q6SQJ9	Q6SQJ9 streptococc
37	1460	12.2	286	2 Q6UK84	Q6UK84 escherichia
38	1460	12.2	286	2 Q6UVM7	Q6UVM7 acinetobact
39	1460	12.2	286	2 Q6WJ74	Q6WJ74 escherichia
40	1460	12.2	286	2 Q7B0V0	Q7B0V0 xanthomonas
41	1460	12.2	286	2 Q9RMS2	Q9RMS2 escherichia
42	1460	12.2	286	2 Q9RN48	Q9RN48 escherichia
43	1459	12.2	286	2 Q93A77	Q93A77 escherichia
44	1459	12.2	286	2 Q9R745	Q9R745 escherichia
45	1458	12.2	286	2 Q48406	Q48406 klebsiella

ALIGNMENTS

RESULT 1
LACI_ECOLI STANDARD; PRT; 360 AA.
ID LACI_ECOLI
AC P03023; O09196; P71309; Q47338;
DT 21-JUL-1986 (Rel. 01, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 25-JAN-2005 (Rel. 46, Last annotation update)
DE Lactose operon repressor.
GN Name=laci; OrderedLocNames=b0345;
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=78246991; PubMed=355891;
RA Farabaugh P.J.;
RT "Sequence of the lacI gene."
RL Nature 274:765-769(1978).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RC MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RC Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,
RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,
RA Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX Chen J., Matthews K.K.S.M.;
RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.
RA Marsh S.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
[5]
RP SEQUENCE OF 1-147; 159-230 AND 233-360.
RX MEDLINE=76091932; PubMed=1107032;
RA Beyreuther K., Adler K., Fanning B., Murray C., Klemm A., Geisler N.;
RT "Amino-acid sequence of lac repressor headpiece from *Escherichia coli*.
RT Isolation, sequence analysis and sequence assembly of tryptic peptides
RT and cyanogen-bromide fragments.";
RL Eur. J. Biochem. 59:491-509(1975).
[6]
RP SEQUENCE OF 1-59; 96-101; 206-215 AND 328-347.
RX MEDLINE=73143730; PubMed=4571224;
RA Platt T., Files J.G., Weber K.;
RT "Lac repressor. Specific proteolytic destruction of the NH 2 -terminal
RT region and loss of the deoxyribonucleic acid-binding activity.";
RL J. Biol. Chem. 248:110-121(1973).
[8]
RP SEQUENCE OF 60-70; 73-78 AND 83-86.
RX MEDLINE=74126378; PubMed=4594037;
RA Ganem D., Miller J.H., Files J.G., Platt T., Weber K.;
RT "Reinitiation of a lac repressor fragment at a condon other than
RT AUG.";
RL Proc. Natl. Acad. Sci. U.S.A. 70:3165-3169(1973).
[9]
RP SEQUENCE OF 1-60 FROM N.A.
RX MEDLINE=88230449; PubMed=3286877;
RA Gordon A.J.E., Burns P.A., Fix D.F., Yatagai F., Allen F.L.,
RA Horsfall M.J., Halliday J.A., Gray J., Bernlot-Moens C.,
RA Glickman B.W.;
RT "Missense mutation in the lacI gene of *Escherichia coli*. Inferences on
RT the structure of the repressor protein.";
RL J. Mol. Biol. 200:239-251(1988).
[10]
RP MUTAGENESIS.
RX MEDLINE=90189956; PubMed=2178920;
RA Lehming N., Sartorius J., Kisters-Woike B., von Wilcken-Bergmann B.,
RA Mueller-Hill B.;
RT "Mutant lac repressors with new specificities hint at rules for
RT protein-DNA recognition.";
RL EMBO J. 9:615-621(1990).
[11]
RP MUTAGENESIS.
RX MEDLINE=94322386; PubMed=8046748;
RA Markiewicz P., Kleina L.G., Cruz C., Ehret S., Miller J.H.;
RT "Genetic studies of the lac repressor. XIV. Analysis of 4000 altered
RT *Escherichia coli* lac repressors reveals essential and non-essential
RT residues, as well as 'spacers' which do not require a specific
RT sequence.";
RL J. Mol. Biol. 240:421-433(1994).
[12]
RP 3D-STRUCTURE MODELING.
RX MEDLINE=91249837; PubMed=2040302;
RA Kisters-Woike B., Lehming N., Sartorius J., von Wilcken-Bergmann B.,
RA Mueller-Hill B.;
RT "A model of the lac repressor-operator complex based on physical and
RT genetic data.";
RL Eur. J. Biochem. 198:411-419(1991).
[13]
RP STRUCTURE BY NMR.
RX MEDLINE=89113344; PubMed=3064080;
RA Boelens R., Lamerichs R.M.J.N., Rullmann J.A.C., van Boom J.H.,
RA Kaptein R.;
RT "The interaction of lac repressor headpiece with its operator: an NMR
RT view.";
RL Protein Seq. Data Anal. 1:487-498(1988).
[14]
RP STRUCTURE BY NMR.
RX MEDLINE=89302886; PubMed=2742823;
RA Lamerichs R.M.J.N., Boelens R., van der Marrel G.A., van Boom J.H.,
RA Kaptein R., Buck F., Fera B., Rueterjans H.;
RT "H NMR study of a complex between the lac repressor headpiece and a 22
RT base pair symmetric lac operator.";
RL Biochemistry 28:2985-2991(1989).
[15]
RP STRUCTURE BY NMR OF 1-56.
RX MEDLINE=96275660; PubMed=8683581; DOI=10.1006/jmbi.1996.0356;
RA Sljaper M., Bonvin A.M., Boelens R., Kaptein R.;
RT "Refined structure of lac repressor headpiece (1-56) determined by
RT relaxation matrix calculations from 2D and 3D NOE data: change of
RT tertiary structure upon binding to the lac operator.";
RL J. Mol. Biol. 259:761-773(1996).
[16]
RP STRUCTURE BY NMR OF 1-62.
RX MEDLINE=20113476; PubMed=10647179; DOI=10.1016/S0969-2126(00)88339-2;
RA Spronk C.A., Bonvin A.M., Radha P.K., Melacini G., Boelens R.,
RA Kaptein R.;
RT "The solution structure of Lac repressor headpiece 62 complexed to a
RT symmetrical lac operator.";
RL Structure 7:1483-1492(1999).
[17]
RP X-RAY CRYSTALLOGRAPHY (4.8 ANGSTROMS).
RX MEDLINE=96239623; PubMed=8638105;
RA Lewis M., Chang G., Horton N.C., Kercher M.A., Pace H.C.,
RA Schumacher M.A., Brennan R.G., Lu P.;
RT "Crystal structure of the lactose operon repressor and its complexes
RT with DNA and inducer.";
RL Science 271:1247-1254(1996).
-!- FUNCTION: Repressor of the lactose operon. Binds lactose as an
CC inducer.
-!- SUBUNIT: Homotetramer.
-!- MISCELLANEOUS: Removal of residues 1-59 and 328-347 results in the
CC molecule losing its DNA-binding activity while maintaining its
CC inducer-binding activity and its tetrameric structure.
-!- SIMILARITY: Contains 1 HTH lacI-type DNA-binding domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
EMBL; V00294; CAA23569.1; -
DR EMBL; J01636; AAA24052.1; -
DR EMBL; U00096; AAC73448.1; -
DR EMBL; U73857; AAB18069.1; ALT_INIT.
DR EMBL; X58469; CAA41383.1; -
DR EMBL; U86347; AAB47270.1; ALT_INIT.
DR EMBL; U72488; AAB36549.1; -
DR EMBL; U78872; AAB37348.1; -
DR EMBL; U78873; AAB37351.1; -
DR EMBL; U78874; AAB37354.1; -
DR PIR; A93198; RPECL.
DR PDB; 1CJG; NMR; A/B=1-62.
DR PDB; 1EFA; X-ray; A/B/C=1-333.
DR PDB; 1JWL; X-ray; A/B/C=1-333.
DR PDB; 1JYE; X-ray; A=1-349.
DR PDB; 1JYF; X-ray; A=1-349.
DR PDB; 1L1M; NMR; A/B=1-62.
DR PDB; 1LBG; X-ray; A/B/C/D=1-359.
DR PDB; 1LBH; X-ray; A/B/C/D=1-359.
DR PDB; 1LBI; X-ray; A/B/C/D=1-359.
DR PDB; 1LCC; NMR; A=1-51.
DR PDB; 1LQC; NMR; A=1-51.
DR PDB; 1LTP; Model; I=1-262.
DR PDB; 1TLF; X-ray; A/B/C/D=60-360.
DR ECO2DBASE; H039.0; 6TH EDITION.
DR EcoGene; EG10525; lacI.
DR InterPro; IPR000843; HTH LacI.
DR InterPro; IPR010982; Lambda_like_DNA.